Data reduction in systems for stereo-sound reproduction using masking effects

Institute for Perception Research
PO Box 513, 5600 MB Eindhoven

28.05.1991

Rapport no. 799

Data Reduction in Systems for Stereo-Sound Reproduction Using Masking Effects

M. del Río González

Data Reduction in Systems for Stereo-Sound Reproduction Using Masking Effects

Report of Practical Work at the
Institute for Perception Research
in Partial Fulfillment of the Requirements for
the Degree of Master of Electronics Engineering
Conferred by the Netherlands University
Foundation for International Cooperation
May 1991

Student:	Manuel del Río González.
Coach:	Dr. A. Kohlrausch /
	Prof. Dr. A.J.M. Houtsma.
Group Leader:	Prof. Dr. A.J.M. Houtsma.
NUFFIC Supervisor:	Prof. Ir C. Kamminga.
EII Supervisor:	Ir P. van der Wurf.

Data Reduction in Systems for Stereo-Sound Reproduction Using Masking Effects

Manuel del Río González

Philips Research Laboratories / Institute for Perception Research

Eindhoven International Institute / Technical University of Eindhoven

Eindhoven, The Netherlands

May 1991

Abstract

The efficient use of masking in subband coding allows the coding of digital-audio signals at very low bit rates without deterioration of the audible quality. Masking is defined as the elevation of the threshold of audibility of one signal or stimulus, the target, because of the presence of another, the masker. In subband coding, the audio signal plays the role of the masker while the quantization noise acts as the target. The use of auditory masking and a structure of filter-banks allows the control of the spectral shape of the quantization noise in such a way that it is efficiently masked by the signal and hence not audible. Then, the bit rate may be reduced and, as long as the signal is able to mask the spectrally-shaped quantization noise, no change in the audible quality of the signal will be perceived.

A basic description of the auditory system is given. A brief introduction to threshold estimation and related topics is included. The conceptual bases behind subband coding are explained. A sequence of experiments designed for the characterization and modelling of the masking of multiple bands of uncorrelated noise (such as quantization noise in subband coding) by a pure tone is described. The theoretical bases are addressed, the experimental procedures implemented are described and the experimental results are presented and discussed from the point of view of finding useful rules, suitable to implementation in subband coding systems.

The first experiment deals with tone-on-tone masking and sets a baseline for the subsequent experiments. The second experiment deals with the masking of single bands of noise of three different bandwidths. A procedure to predict masking data for bands of noise with bandwidths in the range from 1 to 200 Hz is given. The third experiment studies the masking of complex targets composed of two bands of noise and establishes evidence of targets addition; two bands of noise presented simultaneously are less maskable than any of them individually. A model for predicting the reduction in masking observed with the complex targets is proposed. The last experiment explores the summation of multiple targets in conditions resembling subband coding. An 8.2log n rule is found to account for the experimental results. Next, the addition effects on masking are summarized, differentiating additivity of maskers from additivity of targets. Finally, in conclusion, the 8.2log n rule developed is tested by comparison with an example of optimized bit rate reduction performed empirically by trial and error. The rule is found to be consistent.

Acknowledgements

I want to express my gratitude to the Government of Netherlands, to the staff of the Eindhoven International Institute and to all the people working at the Institute for Perception Research (IPO). Thanks to all the people that in some way contributed to make my studies and project work not only interesting but enjoyable.

Contents

1. Introduction 1

2. The Auditory System 2

3. Estimation of Masked Thresholds 6

3.1 Definition of Masked Thresholds 6
3.2 Masking Paradigms 7
3.3 Physiological and Psychophysical Approaches 8
3.4 Threshold Criterion 10

4. Subband Coding 12

5. Tone-on-Tone Masking Experiments 16

5.1 Literature Review 16
5.2 Stimuli and Procedure 18
5.2.1 Signal Definition 18
5.2.2 Apparatus 18
5.2.3 Subjects 19
5.2.4 Procedure 19
5.3 Experimental Results 20
5.4 Discussion 21

6. Masking of a Band of Noise by a Tone 23

6.1 Stimuli and Procedure 23
6.1.1 Signal Definition 23
6.1.2 Apparatus 24
6.1.3 Subjects 24
6.1.4 Procedure 26
6.2 Experimental Results 27
6.3 Discussion 27

7. Masking of Two Uncorrelated Bands of Noise by a Tone 33

7.1 Stimuli and Procedure 33
7.1.1 Signal Definition 33
7.1.2 Apparatus 36
7.1.3 Subject 36
7.1.4 Procedure 36
7.2 Experimental Results 36
7.3 Discussion 37

8. Masking of n Adjacent and Uncorrelated Bands of Noise by a Tone 39

8.1 Stimuli and Procedure 39
8.1.1 Signal Definition 39
8.1.2 Apparatus 40
8.1.3 Subjects 40
8.1.4 Procedure 40
8.2 Experimental Results 40
8.3 Discussion 41

9. Addition Effects in Masking 44

9.1 Additivity of Maskers 44
9.2 Additivity of Targets 46

10. Conclusion 47

10.1 The Use of Additivity Rules in Subband Coding 47

Appendix A Timing of Trials and Signal Smoothing 50
Appendix B Combination Tones 52

Appendix C Software Utilities 54

1. Introduction

Traditionally, audio systems have been designed for the perfect reproduction of waveforms without taking into consideration the psychoacoustic processes performed within the auditory system. Nevertheless, multiple experiments have exhibited limitations of the auditory system with respect to discriminating between signals, suggesting that attempting perfect reproduction of waveforms is not necessary. As long as the difference between the input and the output of an audio system is not audible, some deviation from the original signal may be allowed.

In conventional digital audio systems, the minimum number of bits per sample needed to represent a quantized signal is determined by the signal-to-noise ratio demanded. In principle, 16 bits per sample are necessary to achieve "Compact Disc" quality with a signal-to-noise ratio of more than 90 dB across the audible frequency range.

In *subband coding*, one of the limitations of the auditory system, *auditory masking*, is exploited. Auditory masking results in the elevation of the threshold for the audibility of a signal due to the presence of a stronger signal. When coding, the total bit rate used to represent an audio signal can be reduced as long as the quantization noise is masked by the signal. Proper *spectral shaping* of the quantization noise allows a large bit rate reduction. Masking occurs not only between signals presented simultaneously, but also between signals presented at different times. These types of masking are called *simultaneous* and *non-simultaneous* masking, respectively. Both types of masking could be used in subband coding. The scope of this project is limited to the study of simultaneous masking and its use in subband coding.

2. The Auditory System

The auditory system is the ensemble of organs dedicated to the translation of acoustic signals into perceived sounds. The sounds are preprocessed in the peripheral auditory system and then, processed in the nervous system. The peripheral auditory system is traditionally divided into three subsystems: the external, the middle and the inner ear. The basic functions of these subsystems will be briefly reviewed. The nervous system is not in the scope of this project and it will not be explicitly addressed anymore. For more detailed treatments on the structure and functions of the auditory system, the reader is referred to Green (1976) [1], Rossing (1982) [2] and Scharf and Buus (1986) [3].

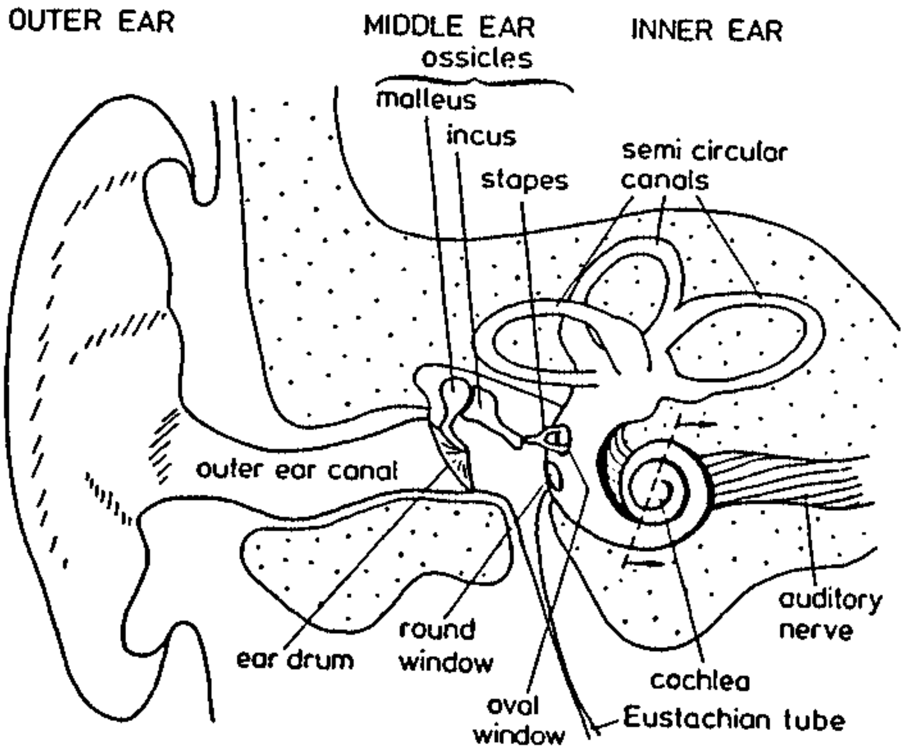


Figure 2.1 *Schematic of the Peripheral Auditory System.*

Even though they are not part of the auditory system, the shoulders and the head are the first parts of our body which affect the perception of a sound field. The reflections from the shoulders and chest, and the acoustic shadow produced by the head, create interference patterns in the acoustic field and modify the reception pattern in a frequency and source-location dependent manner.

The *external ear* consists of the *pinna* and the *external ear canal* (see figure 2.1). The function of these organs is not only protective. The frequency response and the reception pattern of the ear, and hence the localization clues (characteristics used by the auditory system to locate the source of a given sound), are affected by them. The external ear horn-shape serves as an acoustic impedance matcher between the open air environment and the *eardrum* [4,5].

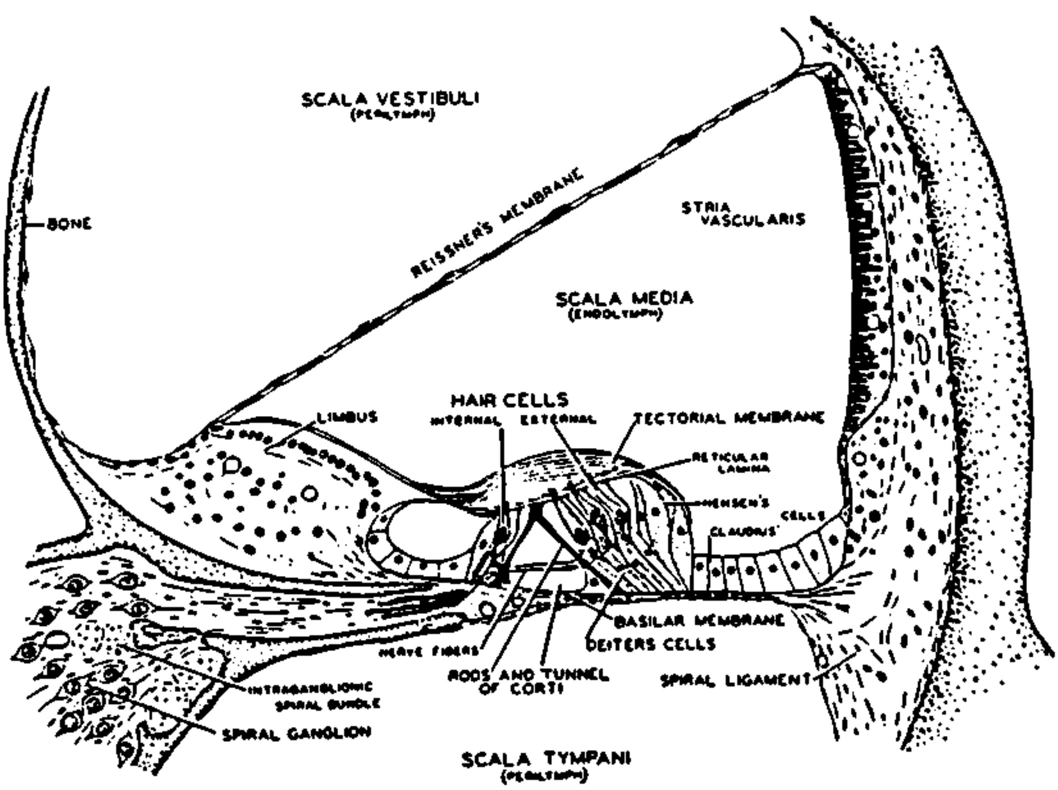


Figure 2.2 *The Cochlear Partition.*

The *middle ear* is composed of the *eardrum* and three *ossicles*. The eardrum is located between the open air environment and the cavity of the middle ear. The acoustic signals are pressure waves which produce a pressure difference between the external and the internal side of the eardrum, which is at constant pressure. The middle ear is located in a closed cavity (the *Eustachian tube* only opens when swallowing). These pressure differences displace the eardrum and make it vibrate. The vibration is transmitted by three ossicles: the *malleus*, the *incus* and the *stapes*. The stapes is connected to the *oval window*, which is the starting point of the *inner ear*. The ossicles work as a mechanical lever system providing impedance coupling between one acoustic medium, the eardrum, and a much higher impedance medium, the fluid inside the *cochlea*. Additionally, the middle ear is able to adapt in such a way as to protect the cochlea against exposure to long and loud sounds.

The *inner ear* comprises the *cochlea* and the *semi-circular canals*. The latter, as far as it is known, perform no function related to hearing. The cochlea is a spiral shaped, fluids-filled cavity which is divided into three subspaces: the *scala vestibuli*, the *scala media* and *the scala tympani* (see figure 2.2). The three scalae run together from the base to the apex of the cochlea. The stapes puts pressure on the *oval window* and displaces it. The displacements of the oval window produce a travelling pressure wave in the fluid filling the scala vestibuli. The scala vestibuli and the scala media are separated only by a very thin membrane, the *Reissner's membrane*. Therefore, the pressure waves in the fluid of the scala vestibuli are also present in the fluid filling the scala media.

The *basilar membrane*, which supports the *organ of Corti*, extends along the *cochlear duct* separating the scala tympani from the scala media. The pressure produced by the travelling wave in the scala media is not immediately compensated in the scala tympani, at the other side of the basilar membrane. This pressure difference produces oscillations of the basilar membrane. At the apex of the cochlear duct, the *helicotrema* connects the scala vestibuli and the scala tympani, allowing the travelling pressure waves in the scala vestibuli to be dissipated towards the fluid in the scala tympani. At the base of the scala tympani, the energy of the travelling waves is used to move the *round window*, allowing the generation of travelling waves in an incompressive fluid that is contained within a bony structure.

The cochlea works as a frequency tuned mechanical system. For a pure tone, the distance from the base of the cochlear duct to the point of maximum displacement of the basilar membrane is dependent on the frequency of the stimulus. Every spot along the basilar membrane responds maximally to only one frequency, thus giving rise to a frequency-place conversion across its length, approx. 35 mm in humans [6,7]. However, each spot responds not only to its "best" frequency but also to nearby frequencies. Greenwood (1990) [8] gives a general function for several species, relating CF (the "best" frequency in Hz) to x (distance in mm from the cochlear base to the place of maximum displacement). Rearranging Greenwood's equation and substituting the proper constants for the human being, we obtain:

$$x = 35 - 16.66 \log[\ (CF/165.4) + 0.88\ ]$$

The *organ of Corti*, located on the basilar membrane, shelters the *sensory cells* in charge of translating the vibrations of the basilar membrane into electric potentials for the nervous system. These sensory cells, the *hair-cells*, are arranged in one row of *inner hair-cells* (IHC) and three rows of *outer hair-cells* (OHC). The IHC are the best suited to pickup the vibrations of the basilar membrane. The OHC, lined in three rows, work as actuators over the basilar membrane as a part of an active, bio-mechanical feedback system responsible for increasing the Q of the mechanical system at low signal levels, and thus improving the frequency selectivity.

The electric potentials arising in the sensory cells are transmitted via the auditory nerve by synapses and similarly, the pathway continues towards the cochlear nucleus, the superior olivary complex, the inferior culliculi and finally the auditory cortex. A remarkable characteristic of the auditory system is its consistent tonotopic organization at the cochlea and at all superior levels, including the cerebral cortex.

3. Estimation of Masked Thresholds

3.1 Definition of Masking

Masking is not only an experimental curiosity, it is found in everyday's life. It manifests itself as the inability to hear a particular sound due to the presence of another louder sound occurring simultaneously or closely in time. A typical situation arises when, at home, a noisy electric appliance is being used and at the same time someone speaks to you. Depending on the intensity of the *masker* (noise from the appliance) and the intensity of the *target* (the person speaking or by now maybe shouting) it can be difficult or even impossible to understand what the speaker is saying.

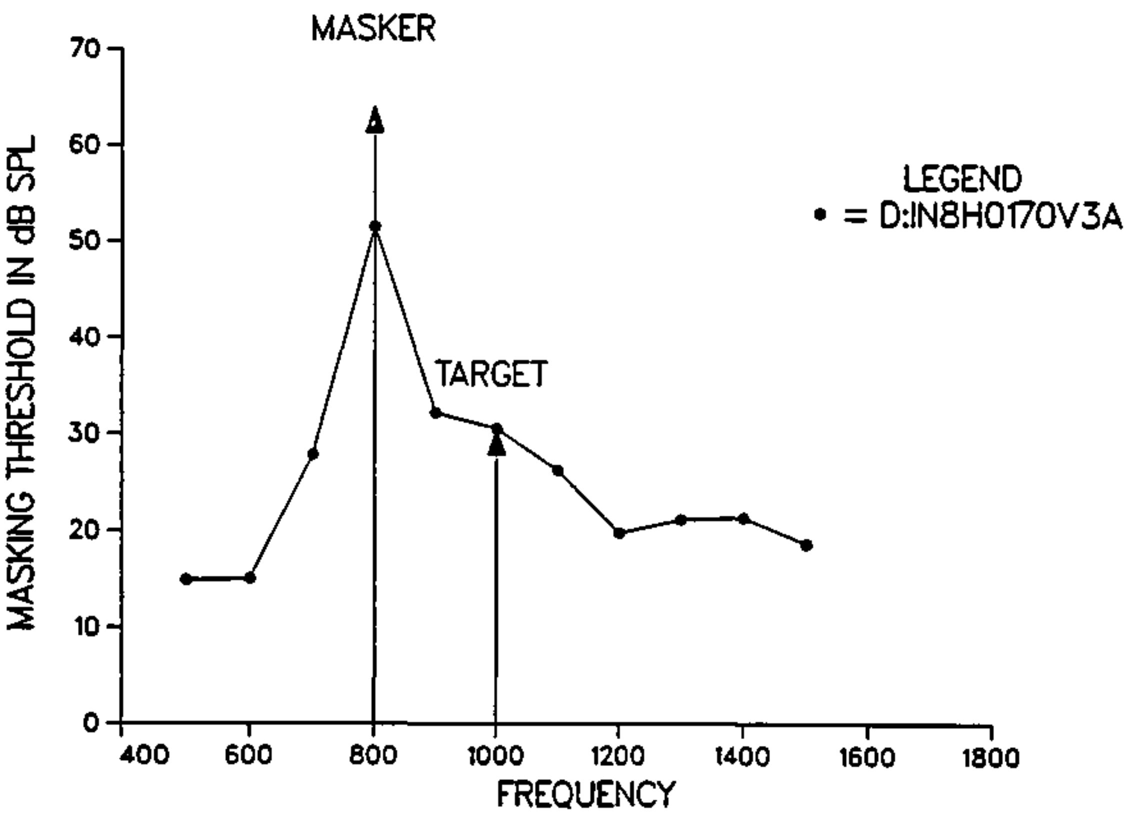


Figure 3.1 *Masking.*

In a more rigorous sense, *masking* is the elevation of the threshold of audibility of one signal or stimulus, the *target*, because of the presence of another, the *masker*. The *masked threshold* is defined as the presentation level necessary for the target signal to be just audible in the presence of the masker. The difference between the target's threshold in quiet and its masked threshold is the *threshold shift* or *amount of masking*.

In relation to the auditory system's physiology, masking may be classified as *peripheral* or *central masking*. Peripheral masking has its origin in the inner ear most probably in the cochlear mechanics (vibration patterns of the basilar membrane). Central masking originates in higher processing stages (neural level). Peripheral masking accounts for most of the measurable threshold shift for long duration signals (more than 200ms). Central masking is particularly efficient for short duration signals (up to 50ms) but significantly reduced for longer duration signals.

3.2 Masking Paradigms

The masker and the target may be presented simultaneously or at different times. Backward, forward or simultaneous masking is measured, depending on whether the target is presented before, after or at the same time as the masker. As mentioned previously, this work is limited to simultaneous masking. For a given combination of masker and target, a masked threshold can be estimated. If the frequencies and/or levels of masker and target are varied systematically, the result of plotting the different thresholds against the value of the chosen parameters is called a masking pattern. A simple situation is illustrated, in figure 3.1. Here, the masking pattern for a pure tone masked by another pure tone of 64 dB SPL (Sound Pressure Level) at 800 Hz is shown. The points indicate the estimated threshold levels when the target was presented at particular frequencies. In this case, the frequency was incremented in steps of 100 Hz starting at 500 Hz and ending at 1500 Hz. In the situation depicted, when the target was presented at 1 kHz, the masked threshold was 32 dB. The maximum threshold was 56 dB SPL when the masker and the target were presented at the same frequency. Observe that the masking extends more towards frequencies above the masker than towards lower frequencies.

When using pure tones for the masker and target, there are four variables that can be used as parameters for a masking pattern: the level of the two tones and the two frequencies. If we are interested only in frequency-level dependencies, there are four possible combinations leading to four different paradigms [6].

1) Input extension pattern: the masker is fixed. The frequency of the target is swept in steps. At each frequency step, the target's level is adjusted until the threshold is found.

2) Input filter pattern: the level of the masker is fixed but the frequency is swept in steps. At each frequency step, the target's level is adjusted until the threshold is found. The target's frequency is fixed.

3) Output extension pattern: the level of the target is fixed but the frequency is swept in steps. At each frequency step, the masker's level is adjusted until the threshold is found. The masker's frequency is fixed.

4) Output filter pattern: the target is fixed. The frequency of the masker is swept in steps. At each frequency step, the masker's level is adjusted until the threshold is found.

As is described in the literature [3,6,10], masking is a non-linear phenomenon. Increasing a masker's level by x dB, does not necessarily increase the masking by x dB (the masking pattern is not just shifted up by x dB). Both filter patterns and the output extension pattern exhibit non-linearity because of presenting the masker at different levels. Besides, both output patterns exhibit larger spread in the results than the input patterns because of the very steep slopes typically found. A small increment in the frequency of the target or the masker leads to a large increment in the level of the masker. For these reasons, input extension patterns are the most common choice. This approach was found to be the most suitable for the experiments performed in the present project.

3.3 Physiological and Psychophysical Approaches

Even though acoustic signals can be uniquely described by the magnitude and phase of their spectral components, the perception of an acoustic signal depends on subjective parameters. The relation between the physical parameters and the percepts (pitch, loudness, timbre, etc.) is not completely known. The auditory system is responsible for the translation of acoustic signals into the perception of sounds. The process undergone in the perception of sounds involves not only physical transformations in the external and middle ear. A great deal of neural processing is performed at the inner ear and higher levels between the eigth nerve and the auditory cortex.

For modelling the auditory system, two main approaches prevail: the physiological approach and the psychophysical approach. In the first, the action of the different organs involved is directly measured. That is, the input-output relationship for each organ in the processing chain is experimentally estimated and modelled. The main idea is to try to understand each part of the process and then to generate a large model which puts all the parts together. The problem is that even the modelling of these "small" parts presents major difficulties, specially at the levels of the neural processing, because too many things remain unknown about its physiology.

The psychophysical approach looks at the whole system as a "black box", concerning itself only with the overall input-output relationship. In the case of the auditory system, the relation between acoustic signal and perceived sound is the input-output function desired. The input-output relation has to be estimated from experimental data. Then, a model able to reproduce the output data from the input can be developed. The model can be based on physiological knowledge. Nevertheless, a psychophysical model is not intended to include the actual processing stages. The specific processing steps occurring inside the "black box" are ignored. The interest is focused on the relation between the input and the output.

Psychophysical modelling has proven to be very useful, but has at least two drawbacks which should be taken into consideration before being too optimistic about its use. Firstly, this kind of models gives little or no detailed insight into the processes involved. Thus, it is very difficult to generalize a result. A model is valid strictly within the assumptions under which it was developed and for the specific inputs and conditions with which the experiments supporting the model were performed. Secondly, the evaluation of the perceived sound relies on the answers from human subjects which are inherently non-deterministic and may be biased. The assessment of reliable measurements of subjective parameters is not a simple task, but it has being widely studied and is well documented.

Masked thresholds can be determined using any of the four paradigms mentioned in section 3.2. A physiological or a psychophysical approach can be selected. In the first, the output from a given organ would be measured with respect to the input applied at a previous stage of the auditory system. The threshold criteria in such an experiment could be fixed in accordance to, for example, the spike rate in an afferent fiber of the eigth nerve (Kiang, 1965) [8]. In the psychophysical approach, the answers from the subject are the output. How the threshold criteria are established in this case, will be treated subsequently.

The psychophysical methods make feasible the study of human responses in cases where physiological experiments would require intrusive techniques such as connecting electrodes to the nerves. The mechanical response of the post-mortem auditory system in humans, exhibits considerable deviations with respect to living subjects. On the other hand, measuring answers from animals is a troublesome task requiring extensive training, if at all possible. The latter makes in many cases the physiological approach more suitable for the study of animal responses. Which method is more suited depends on the specific experiment. In audio systems and subband coding, the relevant issue is the quality of the sound as judged by human listeners. Thus, psychophysical experiments are the best suited to study masking properties useful for encoding audio signals. The experiments on masking described in the following sections fall into this category.

3.4 Threshold Criterion

In order to determine the masked threshold for a given combination of masker and target, several methods may be used. Finding the threshold requires adjusting the level or frequency of the masker or the target in accordance with one of the paradigms given in section 3.2. Subsequently, an input extension pattern paradigm is assumed.

To determine whether or not a target is masked, a response from the subject is needed. It could be left to the subject to adjust a knob, controlling the target level, until finding a level at which the target is just audible. This level could be used to estimate the threshold. Obviously, the subject could bias or even manipulate the results at will. In general, with this method, the subjects underestimate the level needed to detect the target. The magnitude of the bias would certainly depend on the particular subject, introducing a non-controllable variable. These considerations support the need for "tricky" procedures which do not allow the subject to manipulate the results. Another option is using a "yes" or "no" procedure, in which the experimenter presents the target at different levels or do not presents it, and the subject has to tell whether the signal is present or not. Simple in principle, such experiments can be full of flaws produced by a number of human factors. Additionally, when manually performed, these experiments require a long period of time and a large effort from the experimenter and the subjects.

Computer assisted experimental setups allow implementation of sophisticated procedures to change the level of the target and record the responses from the subjects with high accuracy, repeatability and speed. Levitt (1970) [9] presented a technique called "transformed up-down adaptive procedure", which is highly suitable for computer implementation and offers excellent repeatability and speed. For the experiments on masking reported in the following sections, an adaptive procedure was implemented. The particular procedure implemented estimates the threshold as the presentation level needed such that the target is detected in 70.7 % of the trials. An explanation of the implemented procedure and its characteristics are given in section 5.2.4.

4. Subband Coding

Using linear quantization with 16 bits per sample in a straight forward PCM system has become a standard in commercial digital audio. With 16 bits per sample, there are severe limitations due to the large amounts of memory needed and the large bandwidth necessary for transmission of the coded information. The constant struggle to reduce the number of bits per sample without decreasing the subjective signal quality has stimulated the development of more powerful coding techniques.

As explained in section 3.1, a signal of a given frequency increases the threshold of audibility of a second, less intense signal presented in the same frequency region. This elevation of the threshold, called masking, can be used in subband coding as follows:

- The components of a signal which are masked by other components present in the signal, need not to be quantized (that is, they are ignored).

- The quantization noise produced will not be audible if it is masked by the components of the signal.

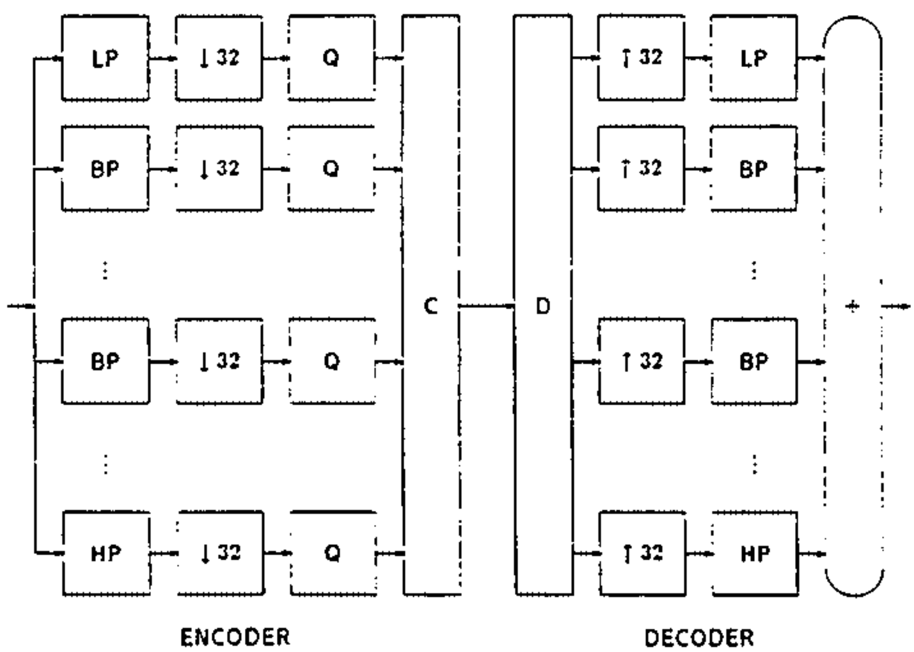


Figure 4.1 *Block Diagram of a Subband Coder*

Subband coding reduces the number of bits per sample needed, by "spending" bits selectively: only in the regions of the spectrum where signal components have to be coded. Components of the signal and quantization noise below the absolute threshold are ignored. In figure 4.1 the schematic of a subband encoder and decoder is shown. In the encoder, an analysis filter bank (boxes LP, BP and HP on the left) divides the audio spectrum into subbands. The number of subbands and their bandwidths is implementation dependent. Once the analysis is performed, the subbands are downsampled. Sampling the signal from each subband at a rate equal to twice its bandwidth is enough to satisfy the Nyquist criterion. At this stage, the number of samples needed for all the subbands is the same as before the analysis and downsampling. In the next step, the quantization of the subbands, is where the bit rate is reduced. In an adaptive bit allocation procedure as described by Veldhuis (1991) [12], two main strategies allow to quantize with fewer bits. First, each band is scaled independently in order to use the full dynamic range of the quantizers and hence shaping the quantization noise to follow the spectrum of the signal. Second, the quantization noise is further shaped in accordance to the estimated masking pattern of the signal. This last point requires further explanation.

Ideally, the noise generated when quantizing one of the subbands is limited at the reconstruction filter bank (boxes LP, BP and HP on the right of figure 4.1) to the spectral width of the subband. Actually, some quantization noise is produced in a given subband by the signal quantized in the adjacent subbands. This is caused by the limited steepness of the filter banks. Nevertheless, by changing the number of bits allocated for the quantization of each subband, the signal-to-noise ratio of each subband can be controlled at will. Using this ability, the signal-to-noise ratio of each subband can be reduced up to the point where the noise is just masked by the signal. With narrow subbands the masking produced by the signal in the nearby frequencies is larger than with wider subbands. Thus, fewer bits need to be allocated for quantizing each subband, but more subbands are required.

In the decoder, the samples from the subbands are recovered after transmission through the channel. The signal of each subband is upsampled by the amount it was downsampled in the coder, and processed by the reconstruction (or synthesis) filter bank. Finally, the signals from all the subbands are merged.

This technique has been explained in a static situation. Music, or audio signals in general, are very dynamic signals. The masking signal and the quantization noise spectra are not static at all. Finding the point in which all the bands of noise are just masked is a complex optimization problem. Algorithms implementing sub-optimal, but very efficient solutions have being developed in several places (e.g. Philips Research Laboratories, The Netherlands). The reduction in bit rates already achieved has encouraged further research in order to get closer to the minimum realizable bit rate without affecting the perceived sound quality. The work reported here is aimed at the study of simultaneous masking, searching for rules which could be useful to predict masking of bands of noise by complex stimuli such as music.

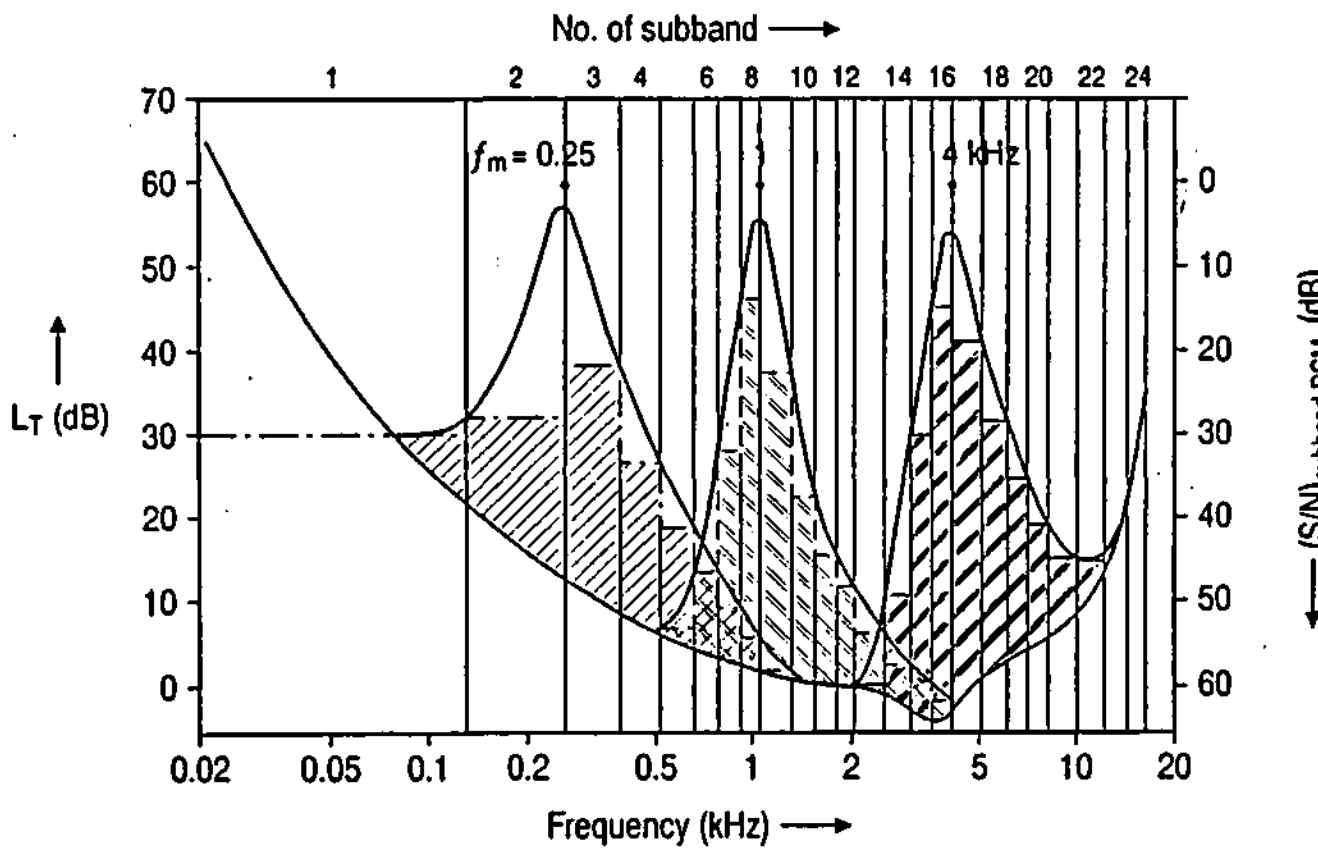


Figure 4.2 *Masking in Subband Coding (descriptive).*

Figure 4.2 depicts a situation in which a complex stimulus with components at 250 Hz, 1 kHz, and 4 kHz is encoded and decoded. Each component produces masking in its surrounding frequency region as represented by the thick, solid lines. The hatched bars represent how the quantization noise in each subband could be allowed to grow as far as it remains below the masked threshold produced by the signal. The picture is only descriptive. If several maskers and several targets are presented together, the masked threshold produced by the complex masker depends on the complex target and it is not just the addition of the masking patterns of the individual maskers over the individual targets. This problem will be addressed in a subsequent section.

5. Tone-on-Tone Masking Experiments

With respect to signals, the masking of a pure tone by another pure tone offers the simplest combination of masker and target. Nevertheless, the analysis of tone-on-tone masking data is complicated by artifacts such as beats and combination tones [4] (see appendix B). In addition, the tone-on-tone masking curves exhibit larger inter-subject variances than those using maskers with a broader spectrum [10]. The idea of characterizing tone-on-tone masking and later to extend the results to complex stimuli by means of Fourier Analysis is appealing. The non-linear response of the auditory system however, limits the usefulness of this approach.

The present experiment was carried out with a limited number of subjects and a reduced number of repetitions. This experiment was not intended to produce extensive data in tone-on-tone masking. Multiple experiments in this topic have been performed before by others and their results are available in the literature [14,15,16,17]. The purpose of this experiment was twofold: to assess the validity of the experimental setup by comparing the data gathered with the data from the literature and to provide a base-line to the following experiments with complex targets.

5.1 Literature Review

As early as 1876, A.E. Mayer demonstrated that low tones could mask high tones, but high tones could not mask low tones [13]. The first extensive study was published in 1924 by Wegel and Lane [13,19]. They found many of the characteristics today known about tone-on-tone masking:

- The masking is largest when the frequency of the target is equal to the frequency of the masker (random-phase tones).
- The masking decreases when the frequency difference between masker and target increases.
- At low levels of the masker, the masking extends only to frequencies very close to the masker. However, at higher levels the masking increases and spreads predominantly towards higher frequencies.

- A small dip found in the masking profile, close to the masker frequency, is attributed to beats. A notch in the masking patterns is found between the frequency of the masker and its first harmonic.

Several authors [14,15,16,17] attribute the "notch phenomenon" in the masking patterns between the masker and its first harmonic to the detection of combination tones (see appendix B). The combination tone at $2f_1-f_2$ is the strongest aural distortion product and therefore the most likely to be detected. The detection is assumed to occur on the lower frequency side of the masker.

Ehmer (1959) [14], measured masking patterns on a wider frequency range, with several masker levels and frequencies. The main observations from this work are the following:

- With a 250 Hz masker, the masking patterns are asymmetrical at all levels while with 500 Hz or higher masker frequencies the patterns are asymmetric only at high masker levels.

- The notch between the masker and the first harmonic shifts upwards in frequency, when the masker level is increased.

- When the masker is below 1 kHz, no notch is observed with masker levels equal to or smaller than 70 dB SPL.

Subsequent experiments performed by Greenwood (1971) [16] supported the findings of Ehmer and added some new observations:

- Incrementing the frequency or the level of the masker produces a gradual transition in the shape of the right side of the masking pattern in the region between the masker and its first harmonic. When the frequency or the level of the masker is increased, the pattern changes from having a fairly continuous slope to having a shelf (or plateau) and finally a notch.

- The frequency interval between the notch's frequency and the masker's frequency increases when the masker frequency is increased. This frequency interval corresponds to a fixed distance in the basilar membrane of 1.25 mm, Greenwood (1971)(or 1 mm, Greenwood (1990)). This distance is very close to the size of the critical band [10](or the EQRB [22]) on the basilar membrane.

- The notch was shown to be present also with bands of noise as maskers and not only with pure tones. This was supported by experimental results which showed that not only combination tones but also combination bands may be generated by the auditory system and hence detected.

- For maskers wider than one critical band and for maskers in which the combination tones and/or the combination bands are masked (the masker used to "cover" the combination products is chosen in such a way as not to modify the region of interest of the masking pattern produced by the original masker), no notches appear in their masking patterns.

- Experimental results support the generation, detection and masking effects of combination tones, combination bands and combination aggregates. Remote masking is also supported.

Finally, Zwicker (1980) [18], extends one of the findings of Wegel and Lane [19] (the third of his findings previously listed). The masking patterns show greater extent towards higher frequencies when the masker level is above 40 dB SPL. Nevertheless, below 40 dB SPL the spread towards frequencies lower than the masker is larger.

5.2 Stimuli and Procedure

5.2.1 Signal Definition

The masker and the target were 16-bits, digitally generated, random-phase sine waves. The masker's frequency (f_M) was fixed at 800 Hz and presented at 64 dB SPL (L_M=64 dB SPL). The target's frequency (f_T) was swept in 100 Hz steps from 500 Hz to 1500 Hz. For each frequency step of the target, the masked thresholds were estimated using the adaptive procedure described in section 5.2.4. The timing of the trials was as given in appendix A.

5.2.2 Apparatus

The experiments were performed under the control of a MicroVAX II computer by means of an IEEE 488 interface. The signals were digitally generated with a DSC 200 D/A convertor system using 16 bits and a sampling frequency of 10 kHz. The signal was low-pass filtered at 4800 Hz with a Kemo Dual Variable Filter. The signals were calibrated and monitored with a DATA 6000 Universal Wave-form Analyzer. The attenuators, signal adders and response recording system were designed and constructed at the Institute for Perception Research, The Netherlands. A signal to noise ratio better than 80 dB was achieved (measured at the input of the headphones). The stimuli were presented *diotically* (same stimulus to both ears) by means of Ethymotic Research ER-2 insert headphones. The experiments were performed in an acoustically isolated environment.

5.2.3 Subjects

Two subjects participated in this experiment. Both subjects were experienced in psychoacoustic experiments. Several familiarization runs were performed before collecting data.

5.2.4 Procedure

The target, as defined in 5.2.1, was swept in frequency steps starting at 500 Hz. First, the threshold for the target was estimated at 500 Hz and then, the frequency of the target was incremented by 100 Hz and the new threshold was estimated. This was repeated until the target reached 1500 Hz. The estimation of the thresholds from 500 Hz to 1500 Hz in the given steps constitutes a single, complete run of this experiment.

In order to estimate the threshold, the subject was asked to indicate in which interval of a three-interval forced-choice procedure [11] the target was presented (the task was explained to the subject in a non-technical and simple way). In one of the three intervals, randomly selected, the masker and the target were presented. In the remaining two intervals, only the masker was presented. The subject had available three buttons that he/she could press in order to express the answer. After answering, the subject was provided with feedback by means of a blinking light. After providing the feedback, a small time was allowed for the subject to assimilate the information and prepare for the next trial. Each trial consisted of three presentation intervals, answer from the subject, feedback and waiting time. The details about timing are given in appendix A.

In the first trial, the target was presented at a level that was known to produce clear detection. This allowed the subject to become familiar with the sound of the target and the buttons that he/she had to press to give answers. Each time the subject gave two consecutive correct answers, the presentation level in the next trial was reduced by 5-dB steps. This continued, making the task more difficult, until the subject failed to give a correct answer. When an incorrect answer was given, the level of the target in the next trial was increased, in steps of 5 dB, making the task easier (entry 2 in [9]). This continued until the subject was able to detect the target or guessed and gave a correct answer. The previous steps were alternatively repeated several times making the presentation level to oscillate around the threshold. Each time the presentation level changed from being increased to being decreased, or vice-versa, it is said that a reversal occurred.

The decision method implemented leads to a threshold value that would elicit detection of the target in 70.7 % of the trials [9]. In order to get an efficient procedure, the step-size for incrementing or decrementing the target's level was adapted [9], starting with steps of 5 dB, 3.25 dB after the first reversal, 2.25 dB after the second, 1.5 dB after the third and 1 dB after the fourth reversal. The number of reversals used to estimate each threshold point was always 15. The threshold level is estimated by averaging the last 10 reversals (reversals number 6 to 15). The standard deviation of the reversals is calculated and can be used as an indication of how consistently the subject is answering. The total number of trials required for the estimation of each threshold point is also recorded.

The forced-choice procedure was implemented with three intervals because this offers better performance than two intervals [11]. Three intervals forced choice procedures offer better efficiency in the estimation (time required to converge to the threshold), less bias and less chance level effects than two-interval procedures [11]. Procedures with four intervals offer a small extra advantage with respect to three-interval procedures. Nevertheless, the improvement is too small to justify the implementation, specially considering that because of having more intervals, the subject may get confused when having to remember the stimulus that he/she heard in each interval.

5.3 Experimental Results

Figure 5.1 shows the average tone-on-tone masking pattern obtained. Each threshold point depicted is the result of averaging four experimental runs: three runs for subject MR and one run for subject IN. The vertical axis corresponds to the threshold of detection for the target and is given in dB SPL. The horizontal axis corresponds to the frequency at which the target was presented. The vertical lines at the threshold points show the standard deviation (n-1 degrees of freedom) of the data from the four runs.

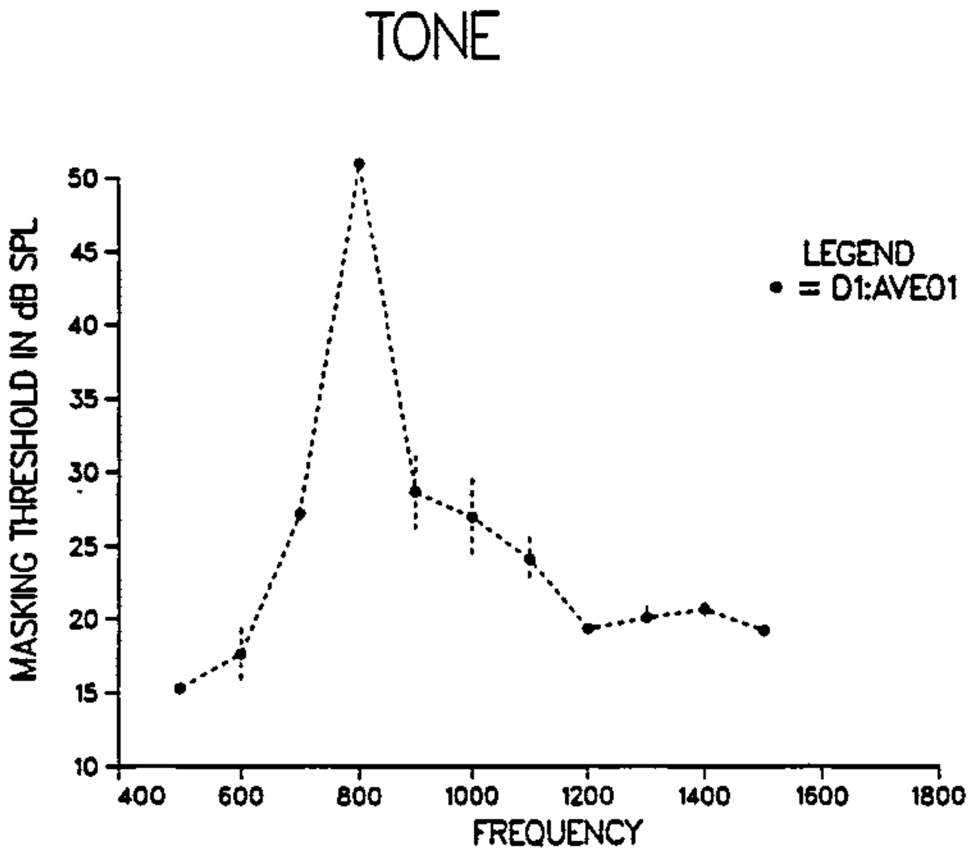


Figure 5.1 *Results of the experiment on Tone-on-Tone Masking.*

5.4 Discussion

Considering figure 5.1, the maximum masking was found to be located at the masker's frequency. Increasing the frequency difference between the masker and the target reduced the masking. The masking was asymmetrical (in linear and in logarithmic frequency scales) and extended more towards frequencies above the masker than towards lower frequencies. The largest masked threshold (for a target presented at the masker's frequency) was found to be 50.5 dB SPL, corresponding to a masker to target ratio of 13.5 dB. Even though it is common custom to fit straight lines to the sides of the masking patterns and to give the slopes obtained as meaningful data, the results presented here do not follow a straight line (nor in a linear nor in a logarithmic axis) making it useless (and impossible!) to give such a slope. Finally, no notch or dip between 800 Hz and 1500 Hz was observed. The frequency step is too large for getting beats, and no evidence of CTs (combination tones) was observed.

The profile and characteristics of the masking pattern are found to be in accordance with the literature reviewed in section 5.1. The masker to target ratio of 13.5 dB is close to 14 dB [16], 15 dB [23], 16 dB [14]. Not finding a notch between 800 Hz and 1500 Hz in the masking data, agrees with the findings of Ehmer [14] (no notch below 70 dB) and Greenwood [16]. The masking data presented here agrees with the literature, supporting the reliability of the results given in the following sections.

6. Masking of a Band of Noise by a Tone

Whereas the masking of tones by noise bands is almost as extensively documented as the masking of tones by tones, the masking of bands of noise by a tone is seldom addressed and no data could be found in current publications. The only aspects on masking of bands of noise that could be predicted from experiments with other kind of targets were:

- Masking should be more efficient when the target's frequency is closer to the masker's frequency. The masking should extend more towards frequencies above the masker than towards lower frequencies.

- For very narrow bands of noise (2-10 Hz), the shape of the masking curves should be very similar to tone-on-tone data but with less beating artifacts. When increasing the bandwidth of the target, the masking curves should show each time smoother transitions and less beating artifacts.

6.1 Stimuli and Procedure

6.1.1 Signal Definition

The masker and the target were digitally generated, using 16 bits. The masker was a random-phase sine wave fixed at 800 Hz (f_M=800 Hz) and presented at 64 dB SPL (L_M=64 dB SPL). The targets were bands of flat-spectrum noise (constant N_0 within the band) taken one at a time. The bands of noise were generated by adding an ensemble of random-phase sine waves, equally spaced (1-Hz interval) within the band's spectrum.

For short-time stimuli and narrow bandwidths, the energy of a noise band and its peak factor (time structure or envelope) have a large variance. In order to avoid artifacts produced by stimuli (bands of noise) with non accurately controlled energy (SPL) or with a particular time structure, two measures were taken:

- First, the stimuli consisted of "pseudo-running" noise and not "frozen" noise. The "pseudo-running" noise was achieved by generating long-duration noise bands and taking from them a new, random, short-duration sample each time the stimuli had to be presented. This method avoids presenting always the same noise band which could posses a particular temporal structure making it especially detectable or especially undetectable [20]. When many different noise bands, each one with a different time structure, are used, the differences are averaged and the estimation procedure converge.

- Second, each time a random sample was selected from the long-duration noise band signal, and after putting on the attack and decay (see appendix A), the RMS (Root Mean Squared) value of the samples was calculated. With the RMS value of the samples, a scale factor was calculated and the stimulus was scaled to have always the same RMS value. For a 0-mean signal (as the ones used), a constant RMS value implies constant energy. Having noise bands with a constant energy allowed an accurate control of the stimuli's SPL.

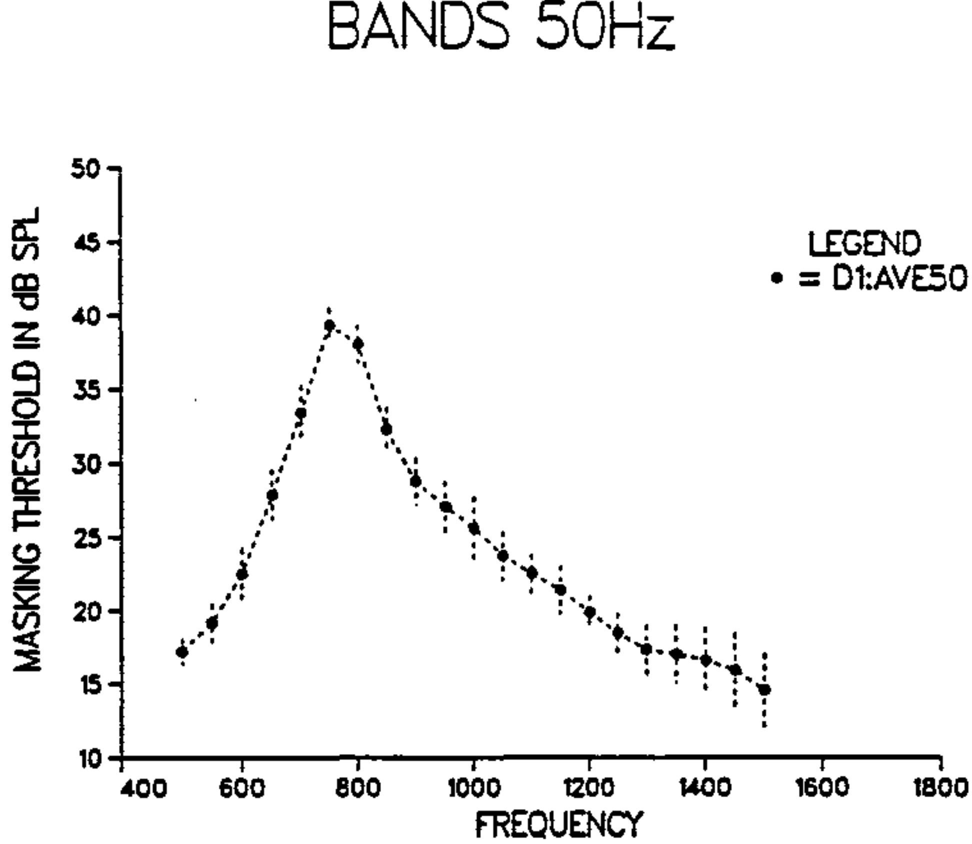
The targets were bands of noise, as described above, with bandwidths (BWs) of 50 Hz, 100 Hz and 200 Hz. The bands were non-overlapping, covering the frequency range from 500 Hz to 1.5 kHz. Twenty-one bands of 50 Hz (500-550,550-600,600-650,etc.), eleven bands of 100 Hz (500-600,600-700,etc.) and six of 200 Hz (500-700,700-900,etc.). The reference of the bands was fixed at the low-frequency edge (not the center!). In the masking paradigm, the noise band was swept in frequency steps from 500 Hz to 1500 Hz. The frequency steps were equal to the bandwidth the target. The masking patterns for the bands of noise of the three BWs were estimated independently. The method to estimate the masked thresholds and the timing of the signals was the same as for the tone-on-tone masking experiment (section 5).

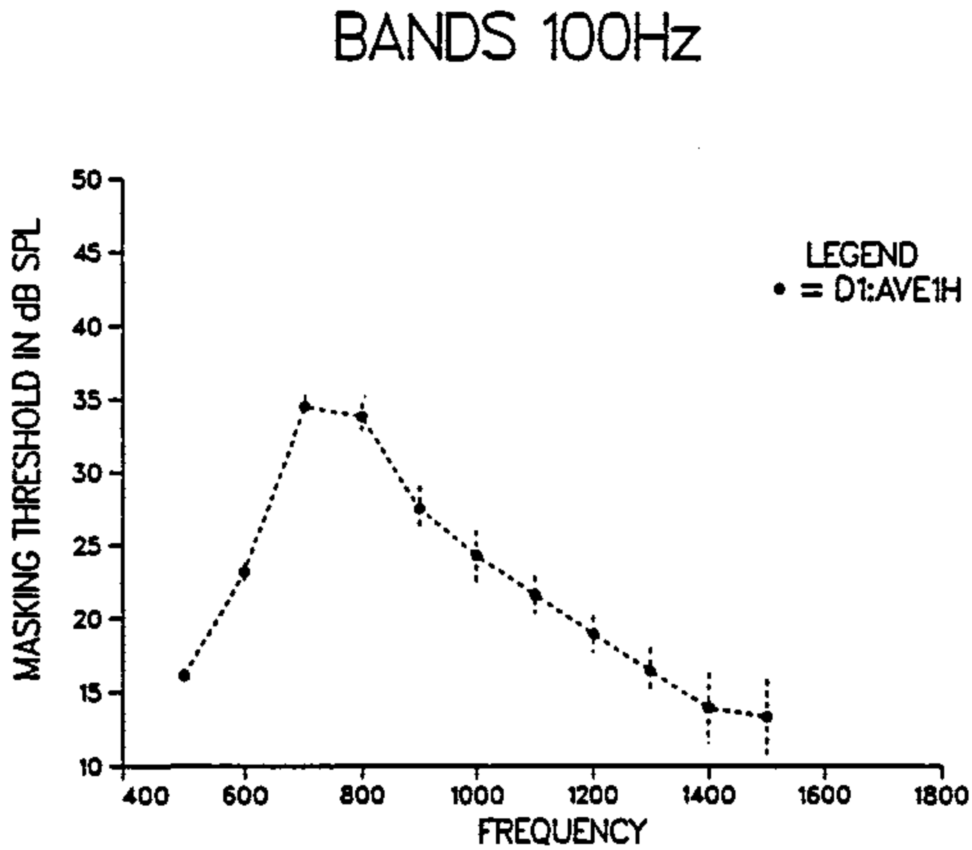
6.1.2 Apparatus

The setup was the same as for the tone-on-tone masking experiment (see section 5.2.2). A signal-to-noise ratio better than 70 dB was achieved (measured at the input of the headphones).

6.1.3 Subjects

Five subjects participated in this experiment. Three of them were experienced in psychoacoustic experiments. The two subjects who took part in the tone-on-tone masking experiment also took part in this experiment. For each subject, several familiarization runs were performed before collecting data.





Figures 6.1 and 6.2 *Results of the experiments on masking of a single band of noise for BW=50Hz and BW=100Hz.*

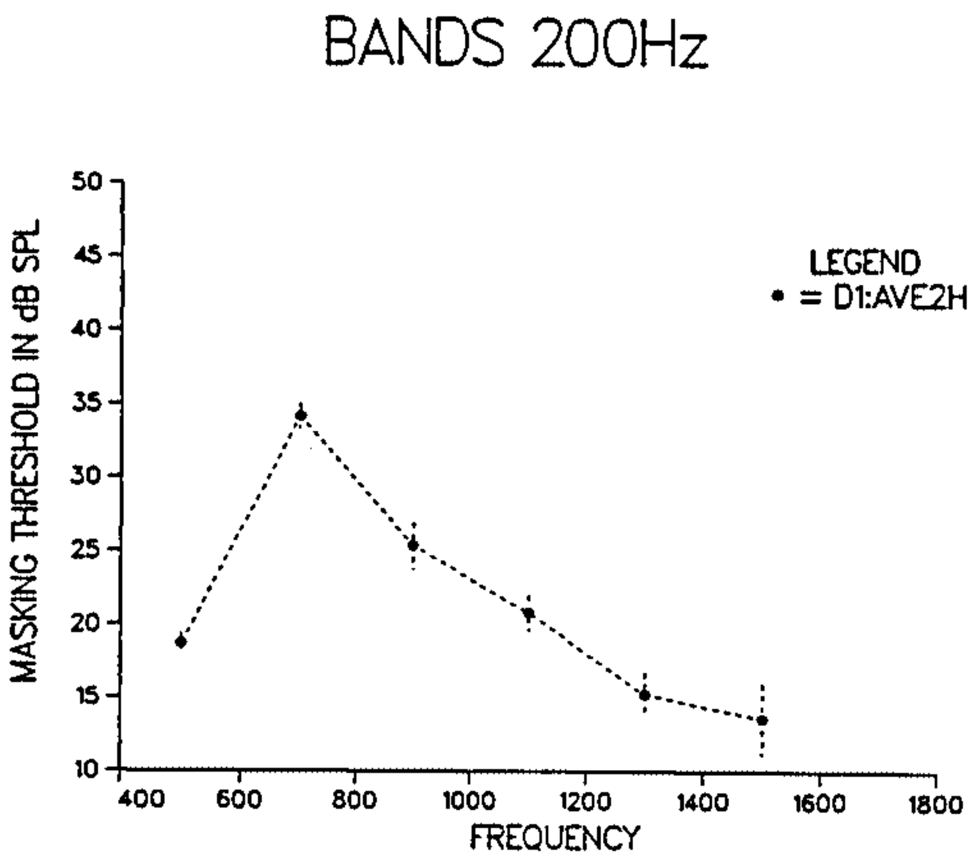


Figure 6.3 *Results of the experiment on masking of a single band of noise for BW=200Hz.*

6.1.4 Procedure

The same procedure used for the tone-on-tone masking experiment (see section 5.2.4) was applied for this experiment. The only difference was that in this experiment, a single, complete run was carried out for each one of the three bandwidths.

6.2 Experimental Results

Figures 6.1, 6.2 and 6.3 show the average masking patterns obtained with bands of noise of 50 Hz, 100 Hz and 200 Hz respectively. Each threshold point depicted is the result of averaging a minimum of 15 experimental runs. Three to five runs were performed with each one of the 5 subjects (for bands of 100 Hz and 200 Hz only the data from 4 subjects were averaged). The average was calculated first between the several runs for each subject (subject average) and later between the subjects (inter-subject average). The vertical axis corresponds to the threshold of detection for the target and is given in dB SPL. The horizontal axis corresponds to the frequency at which the low-frequency edge of the noise band was presented. The vertical lines at the threshold points show the standard deviation (n-1 degrees of freedom) of the data averaged for each threshold point. Figure 6.4 shows the masking patterns for the three widths of noise bands plotted together. The same axes as in figures 6.1-3 were used.

6.3 Discussion

The masking patterns for noise bands (figures 6.1-3) showed smoother transitions than the masking pattern for tone-on-tone (figure 5.1). Widening the spectrum of the target resulted in smoother patterns because the fine details are averaged across the frequency width of the target. Increasing the width of the target shifted the peak of the masking patterns towards lower frequencies because the reference of the bands was taken at their left frequency edge. This means that in order to make the target as difficult to detect as possible, it had to be centered below the excitation pattern [22] of the masker. As expected, the masking extends more towards frequencies above the masker's frequency than towards those below. On a logarithmic frequency scale the three masking curves exhibit similar slopes between 14 and 30 dB SPL of masked threshold. 53 dB/octave in the low-frequency side and 22 dB/octave in the high-frequency side. At higher levels, the curves for narrow bands show steeper slopes. The maximum masked threshold (peak) decreased each time the target's bandwidth was increased.

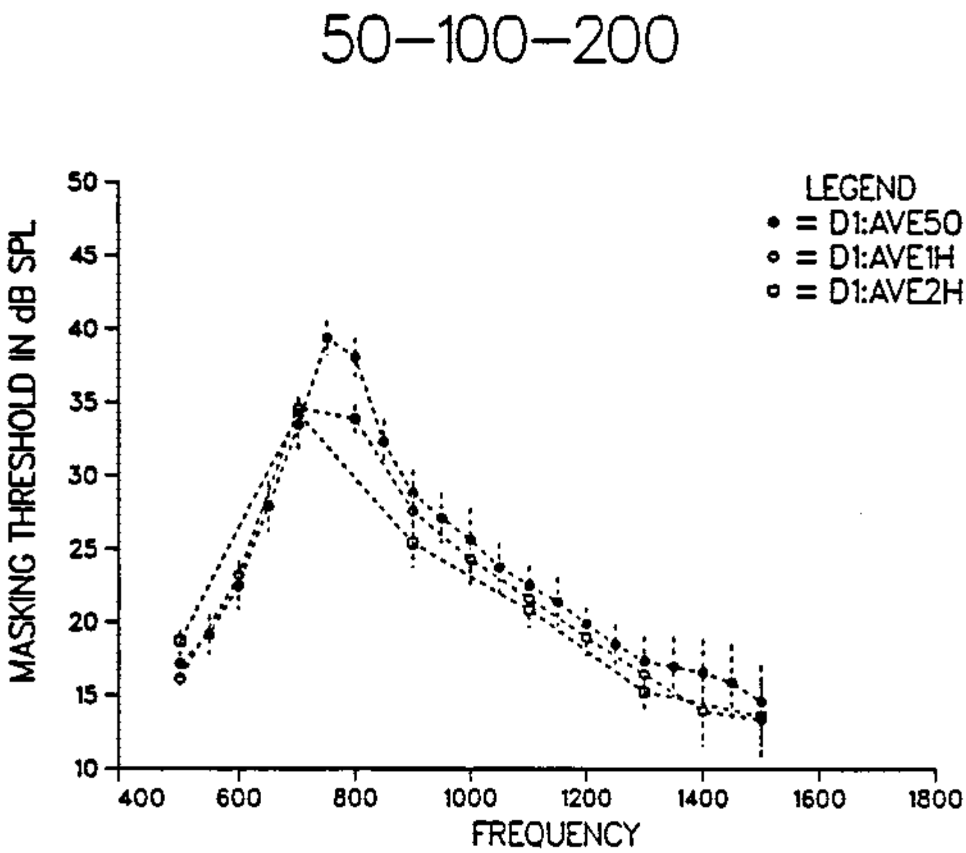


Figure 6.4 *Results of the experiment on masking of a single band of noise for BW=50Hz, BW=100Hz and BW=200Hz plotted together, the frequency reference of the bands is taken in the lower-frequency edge.*

Increasing the bandwidth of the targets while keeping the energy of the target constant requires a proportional reduction of the spectral density level (N_0). Each time the bandwidth is doubled, N_0 has to be halved, as shown by the equation:

$$\text{Band Energy} = N_0 \; BW$$

In figure 6.4, the thresholds increased slightly each time the bands were widened, in the region of the low-frequency slope. Because the reference of the bands was taken at the low-frequency edge, widening a band places components of the target closer to the frequency of the masker (where the masking is most efficient) and hence this energy is not efficiently used for detection. This explanation could account for an increase of up to 3 dB in the threshold of a band when its bandwidth is doubled. This maximum limit of 3 dB would be reached only if none of the energy placed at higher frequencies were to be used by the detection mechanism (a masker with an excitation pattern infinitely steep). In this extreme case, the double-width target would not be detected until the N_0 is doubled thus reaching the level of the half-width target with a total energy increase of 3 dB. In this experiment, the slope of the masker's excitation pattern was not so steep (only 60 to 80 dB/octave) and an increase of 1-2 dB in the target's energy was enough to reach threshold after doubling the target's bandwidth.

On the other hand, at the high-frequency slope of the patterns a decrease in masking threshold was observed. When the target was widened, their components were placed in a frequency region each time farther away from the masker's frequency thus providing a better target-to-masker ratio for detection. Again, the reduction in the target's energy needed for detection depends in the slope of the masker's excitation pattern. The steeper the slope, the larger the reduction in the energy required. If perfect across-frequency integration of the energy were to occur, the target's threshold would remain constant for a masker with flat excitation pattern independently of its bandwidth.

In the region of the peak ($f_M \pm$ BW/2), the target may go out of the maskers excitation pattern in any of both frequency directions. Qualitatively, the wider the masker, the less it will be able to fit inside the narrowing masker's excitation pattern before going out in one of the sides or both and being detected.

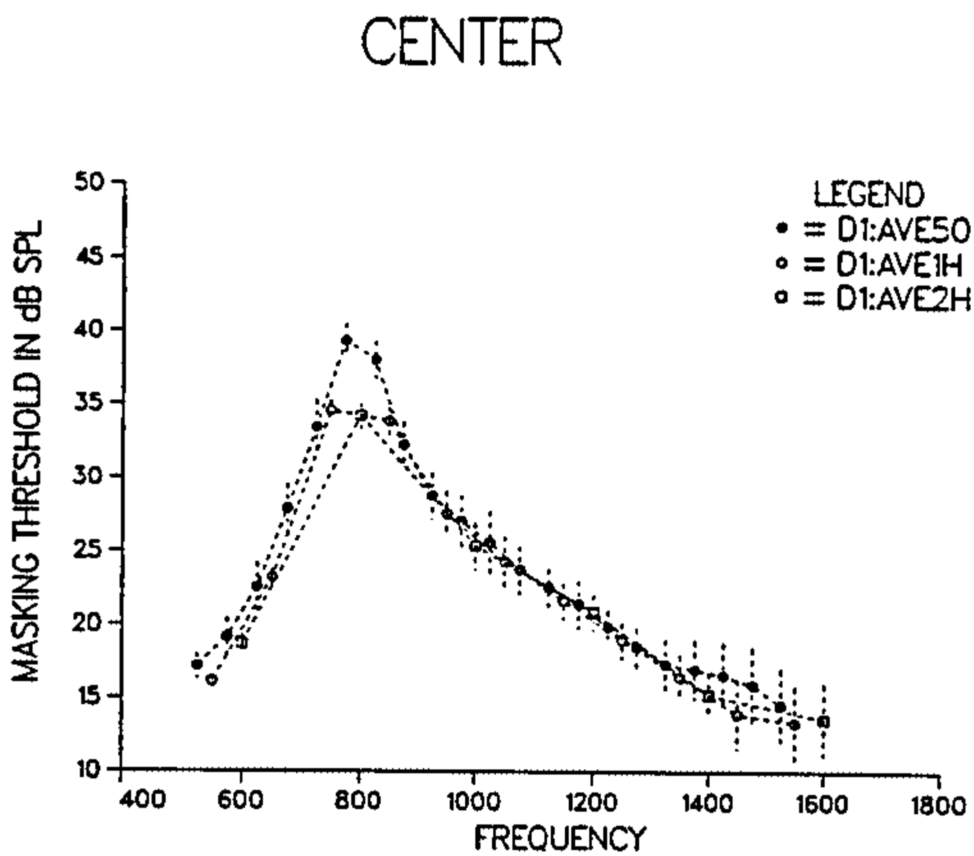


Figure 6.5 *Results of the experiment on masking of a single band of noise for BW=50Hz, BW=100Hz and BW=200Hz plotted together, the frequency reference of the bands is taken in the center frequency of the bands.*

For further comparisons, the three masking patterns were plotted as in figure 6.5, taking as frequency reference of the targets the arithmetic center of the bands. In this figure, the three masking patterns were close to match each other. The reason for the differences in the region of the peak has already been addressed. Why the match was better at the high-frequency slope than at the low-frequency slope, is not completely clear. Two hypotheses are proposed:

1) The slope of the masker's excitation pattern is steeper in the low-frequency side than in the high-frequency side. Thus, a larger change in the masked threshold can be expected than in the high-frequency side which is less sensitive to changes in the target's bandwidth.

2) The combination aggregate (see appendix B) at the lower edge of a noise band, becomes stronger when its bandwidth is increased. These aggregates may help for the detection of wider noise bands.

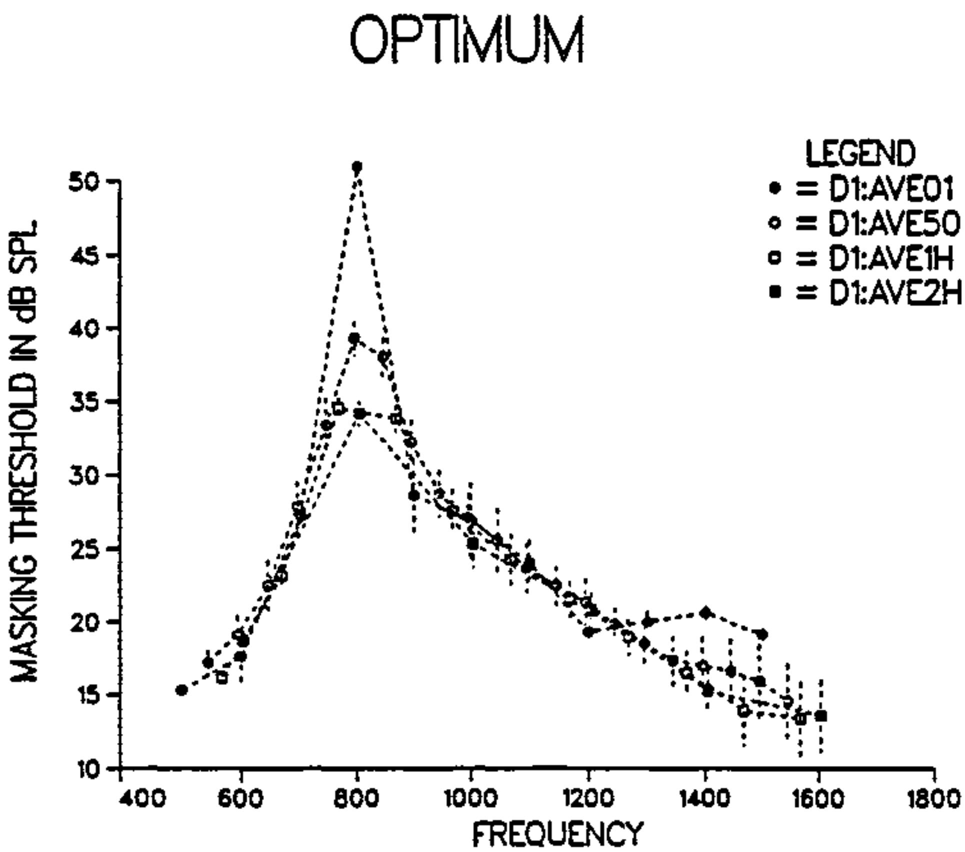


Figure 6.6 *Results of the experiment on masking of a single band of noise for BW=50Hz, BW=100Hz and BW=200Hz plotted together, the frequency reference of the bands is taken in the "effective frequency" of the bands.*

If the masking patterns for the different bandwidths were matched in some way, a tool for predicting masking patterns of noise bands by tones would be available. In order to obtain a better match than in figure 6.5, the masking patterns were shifted on a linear frequency axis until finding the target's frequency reference that produces a best fit. The idea is that if the masking pattern for a noise band in the range studied here is taken as the target's frequency reference (the so called *effective frequency* of the target), a very similar pattern (except for the peak region) would be obtained. Figure 6.6 shows the result of plotting the masking data from these experiments with respect to the *effective frequency* of the targets. The tone-on-tone data is included as a reference. The *effective frequency* (f_{eff}) was calculated in accordance to an equation fitted to match the data:

$$f_{eff} = [(f_h+f_l) / 2] + [1 - 0.55(10^{(BW/800)})] [BW-1]$$

where f_l and f_h are the low and high frequency edges of the band, in Hz. The first term corresponds to the arithmetic center frequency of the band. The f_{eff} can be used to approximate the masking patterns of noise bands by tones following the next steps.

1) First, assume a masker at 800 Hz. For each target of given BW, f_l, and f_h calculate the f_{eff}. Using the f_{eff} and figure 6.6 (slopes of 53 and 22 dB/octave) get the corresponding masked threshold. Do this for all the target frequencies.

2) Using the fact that the shape of the masking patterns is quite independent of frequency in a EQRB or a critical-band scale, translate the pattern just obtained to the corresponding masker's frequency. The peak should be located at f_M.

3) The curve obtained should give a good approximation away from the peak region ($f_M \pm BW/2$), but the level of the peak must be changed to:

$$\text{Peak dB} = L_M - 10.5 + a_0 - 7.8 \log BW$$

and the slope in the $f_M \pm BW/2$ region can be approximated by a line between the peak and the level of the skirts given in 1) at the points $f_M \pm BW/2$. The term a_0 is to compensate for different detection sensibility at different frequencies in accordance to Zwicker [10].

7. Masking of Two Uncorrelated Bands of Noise by a Tone

As for the masking of single bands of noise by a tone, no literature could be found about masking of two or more bands of noise by a tone. A serious limitation of this experiment was that there were many free parameters: the bandwidth of each band (BW), the frequency difference between the two bands (ΔB), the frequency region to perform the experiment and the level difference between the two bands (ΔL). Furthermore, even in the case that the influence of all the parameters were estimated experimentally, the usefulness of the results would be limited to very simple targets. Nevertheless, the selection of some particular combinations of parameters produced interesting results. From the experiment reported in the previous chapter and applying some detection theory, some aspects about the masking of two bands of noise could be predicted.

- Assuming that any mutual suppression effect is negligible, then if any of the two bands composing the target is detectable by itself, the compound target should be detectable.

- If a compound target is formed by adding two bands of noise with equal frequency, bandwidth and N_0, the threshold for the combined target should be 3 dB below (more detectable) the threshold for one of the bands alone (remember they are uncorrelated).

- When the energy of one of the bands is considerably larger than the energy of the other, the larger band dominates and the threshold for the combined target should be very close to the threshold of the stronger target alone.

7.1 Stimuli and Procedure

7.1.1 Signal Definition

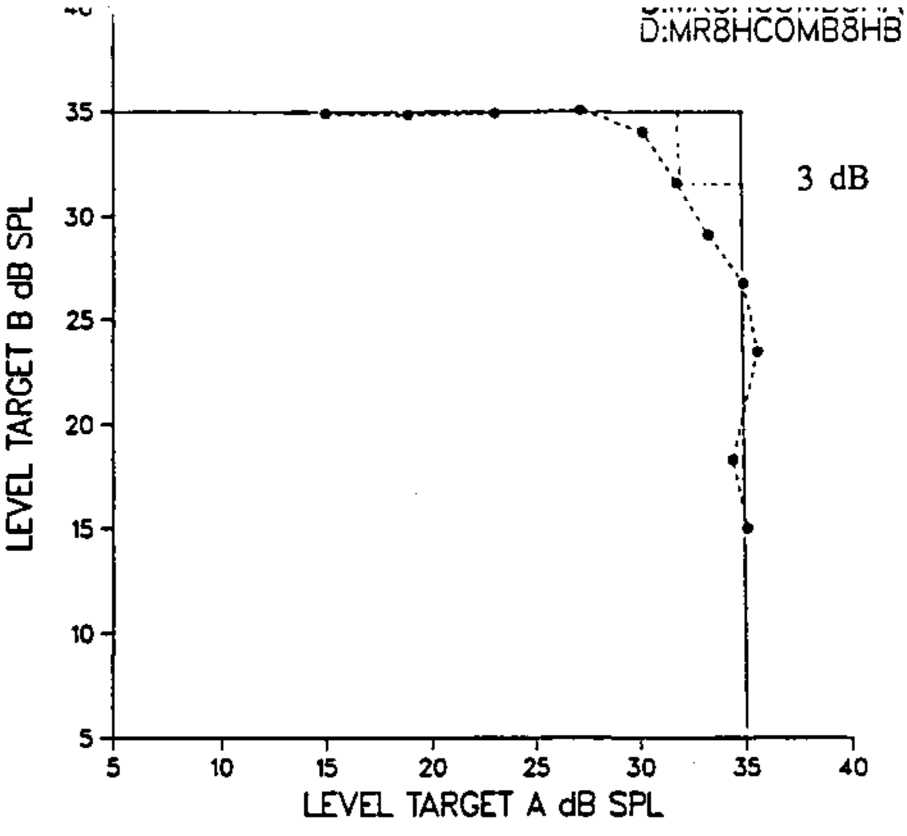
The masker and the target were digitally generated, using 16 bits. The masker was a random-phase sine wave fixed at 800 Hz and presented at 64 dB SPL. The combined targets were generated by adding two bands of flat-spectrum noise as described in section 6.1.1. In order to prevent correlation, the two bands were generated independently and then added. The energy of each band was scaled (see section 6.1.1), in order to produce the desired ΔL, before performing the addition. The same technique for producing "pseudo-running" noise, described in section 6.1.1, was again used.

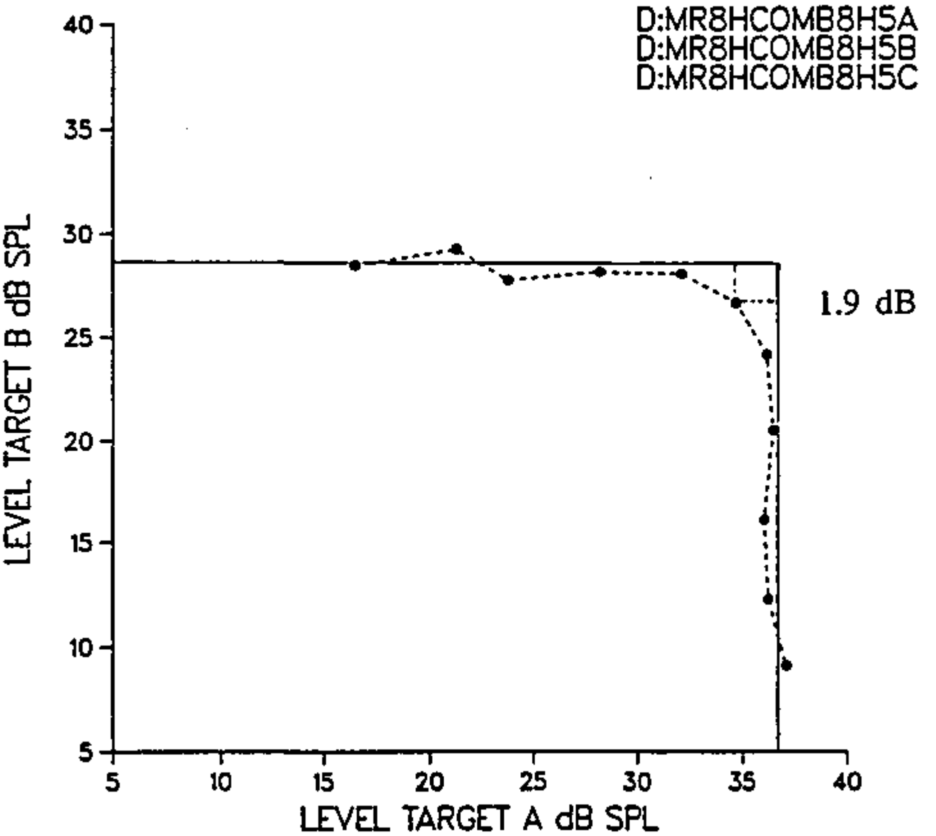
Four stimuli (combined targets) were generated by adding bands of noise with BW = 50 Hz and lower-edge frequencies already studied in chapter 7. These allowed for comparison between the thresholds for each band alone and the threshold for the combined target. Each pair of bands was presented with 11 inter-band level differences (ΔLs) systematically incremented. The ΔLs for each pair of bands were calculated in such a way that the 6^{th} made both bands equally detectable. Hence, in principle, with the first five ΔL, one of the bands was more detectable and with the last five ΔL, the other band was more detectable. This was done in accordance with the data on the masked thresholds for single bands already gathered (see chapter 6). Two bands become equally detectable if the difference between the masked threshold of each one and the presentation level of each one are equal (i.e. band A is presented 10 dB below its threshold and band B is also presented 10 dB below its threshold). This definition of equal detectability is only valid for stimuli with a small number of components. For stimuli with more components, the extent of the excitation of the components with lower frequencies adds to the excitation of the components at higher frequencies and makes them more detectable. Hence, the equal detectability assumption would not be valid any more (for more on this see [21]). The two bands combined to form a compound stimulus were called "Target A" and "Target B". The four combinations of frequencies were as follows:

Combination #	Lower-edge Freq. Target A	Lower-edge Freq. Target B
1	800 Hz	800 Hz
2	800 Hz	850 Hz
3	1000 Hz	1050 Hz
4	700 Hz	850 Hz

Note: The frequencies of the targets are given with reference to the lower-frequency edge of the bands.

The method used to estimate the masked thresholds and the timing of the signal was the same used for the experiments described in chapters 5 and 6.





Figures 7.1 and 7.2 *Results of the experiments on masking of two bands of noise with the combinations of targets 1 and 2.*

7.1.2 Apparatus

The setup was the same as for the tone-on-tone and for the single band masking experiments (see section 5.2.2). A signal to noise ratio better than 70 dB was achieved (measured at the input of the headphones).

7.1.3 Subject

Only one subject participated in this experiment. The subject was experienced in psychoacoustic experiments. Several familiarization runs were performed before collecting data.

7.1.4 Procedure

The same procedure used for the tone-on-tone masking experiment (see section 5.2.4) was again applied for this experiment. The only difference was that in this experiment, a single, complete run consisted of estimating the threshold for a given combination of bands with the 11 different ΔLs.

7.2 Experimental Results

Figures 7.1 and 7.2 show the results for the combinations of targets number 1 and 2. Similar results were obtained for the combinations 3 and 4 but they are not shown. Each threshold point was the result of averaging 3 runs (only 2 runs for combination 1). For a given threshold point, the value in dB SPL in the horizontal and vertical axes correspond with the levels of target A and target B that combined produced a just detectable stimulus (at threshold level).

The following table gives the maximum summation of the targets found for each frequency combination.

Combination #	Freq. Target A	Freq. Target B	Max. Addition
1	800 Hz	800 Hz	3 dB
2	800 Hz	850 Hz	1.9 dB
3	1000 Hz	1050 Hz	2.3 dB
4	700 Hz	850 Hz	1.9 dB

Note: The frequencies of the targets are given with reference to the low frequency-edge of the bands.

7.3 Discussion

From the figures and the previous table, there is evidence of summation between the targets. When two bands were presented simultaneously, each one of them at masking threshold, they were clearly audible. In order to reach the masked threshold again, the presentation level of the combined target had to be reduced. Then, presenting each target at a level far below their individual thresholds, the two targets combined were able to reach threshold. It can be observed in the figures that when one of the targets is considerably stronger than the other ($\Delta L > 15$ dB), the threshold is determined exclusively by the stronger one (dominant). This was supported by the masking data from chapter 6. The ΔL which produced the maximum summation gives information about the detection process. This experiment supports the idea that two bands of noise cooperate most efficiently for detection, when they are equally detectable.

The auditory system works as a stochastic detector and can well be analyzed as such. With the threshold criterion given in 3.4, the probability of detecting a band of noise "A" at threshold level is 0.7071. If a second band "B" is simultaneously presented at its threshold level, and independent detection is assumed, the probability of detecting the compound stimulus is:

$$P(A=1) = P(B=1) = 0.7071$$

$$P(A+B=1) = 1 - P(A=0,B=0) = 1 - (0.7071)^2 = 0.9142$$

It is clear that even assuming independent detection, a sort of summation between the targets occurs and makes the combined target more detectable. If the combined target were to be driven to threshold, its level would have to be reduced. To calculate the amount of reduction necessary, the psychometric function could be used. In general, assuming independent detection , with "p" equal to the probability of detection with a given threshold criterion, if $P(C=1)$ is the probability of detecting the combination of n equally detectable stimuli,

$$P(C=1) = 1 - (1-p)^n$$

In this experiment, for the combination of two equally detectable targets, it was found that the reduction in level necessary to drive the combined target to threshold was 2 dB in average. This is half way between the prediction from a statistical summation model of 1.5 dB (5log 2) [21] and perfect integration (3 dB or 10log 2). This result is interesting but not directly comparable because of the following reasons:

- The 5 log n rule was derived from thresholds producing detection in 79.4 % of the trials and not 70.71 % as in this experiment.

- The condition of equal detection is approximate in this experiment (because of the small number of bands), but no special provisions were taken to ensure independent detection of the targets.

In order to test the validity of the statistical model to predict masking of compound stimuli, care must be taken on the assumptions concerning independent detection and equal detectability. The excitation pattern of one stimulus extends and interacts with those of other stimuli, affecting the detectability of the whole. Furthermore, to distinguish whether or not two stimuli belong to different detection channels is not obvious. Zwicker and Feldtkeller (1967) proposed to divide the stimuli in 24 independent channels (similar to critical bands [10]) across the audible range.

8. Masking of n Adjacent and Uncorrelated Bands of Noise by a Tone.

After studying the masking of single bands of noise by a tone and finding that when two bands of uncorrelated noise are presented simultaneously they cooperate for detection, the next logical step is the study of multiple bands of noise as targets. The conditions for this experiment were established in such a way that the results could be suited for implementation in subband coding. The conditions follow:

1) The bands were ucorrelated.

2) All bands were adjacent (no spectral gaps) and had equal bandwidth.

3) All bands were an equal number of decibels below the masked threshold of the individual bands. This produces something close to equal detectability, but not exactly equal detectability of the individual targets (see section 7.1.1).

8.1 Stimuli and Procedure

8.1.1 Signal Definition

The masker and the target were digitally generated, using 16 bits. The masker was a random-phase sine wave fixed at 800 Hz (f_M=800 Hz) and presented at 64 dB SPL (L_M=64 dB SPL). The combined targets were generated by adding a set of 6, 11 or 21 bands of flat-spectrum noise as given in 6.1.1. In order to prevent correlation, all the bands were generated independently and then added. The energy of each band was scaled (see section 6.1.1), in order to produce the desired close-to-equal detectability, before performing the addition. The same technique for producing "pseudo-running" noise, described in 6.1.1, was used.

Three stimuli (combined targets) were generated by adding 6, 11, and 21 bands of noise of 200, 100 and 50 Hz BW respectively. The three stimuli were as follows:

Stimulus	Numb. of Bands	W per Band	Total Frequency Range
1	6	200 Hz	500-1700 Hz
2	11	100 Hz	500-1600 Hz
3	21	50 Hz	500-1550 Hz

The method to estimate the masked thresholds and the timing of the signal was the same as used for the experiments described in chapters 5, 6 and 7.

8.1.2 Apparatus

The setup was the same as for the other experiments (see section 5.2.2). A signal-to-noise ratio better than 70 dB was achieved (measured at the input of the headphones).

8.1.3 Subjects

Two subjects participated in this experiment. Both subjects were experienced in psychoacoustic experiments.

8.1.4 Procedure

The same procedure used for the tone-on-tone masking experiment (see section 5.2.4) was again applied for this experiment. The only difference was that in this experiment, a single, complete run consisted of estimating three times the threshold for a given combination of bands.

8.2 Experimental Results

The results of this experiment are summarized in the following table. The value for two bands of noise was taken from the experiment with two bands of noise in chapter 7. The addition value indicates the attenuation that had to be given to the combined target in order to reach the threshold again.

Stimulus	Numb. of Bands	Addition	Total Frequency Range
-	2	2.1 dB	(see note)
1	6	6.5 dB	500-1700 Hz
2	11	8.7 dB	500-1600 Hz
3	21	12.2 dB	500-1550 Hz

Note: For the stimulus with two bands, the addition value is the average addition for the bands studied in chapter 7.

8.3 Discussion

It is evident that strong addition occurs and as much as 12 dB of attenuation was required to decrease the compound stimuli to threshold, even though the components were presented at their individual threshold level. The shift in the thresholds is not far from that which could be expected assuming perfect across-frequency integration of the energy of all components. The following table gives the experimental shifts, the expected shifts from a 10log n rule and the results calculated using a rule fitted to the data.

Stimulus	Number of Bands	Experimental	Addition 10log n	8.2log n
-	1	0 dB	0.0 dB	0.0 dB
-	2	2.1 dB	3.0 dB	2.4 dB
1	6	6.5 dB	7.8 dB	6.4 dB
2	11	8.7 dB	10.4 dB	8.5 dB
3	21	12.2 dB	13.2 dB	10.8dB

The fact that an 8.2log n rule is able to account for the results of this experiment suggests that the across-frequency integration for uncorrelated noise bands as targets under the given conditions is rather efficient. This means that efficient integration may hold also for larger duration signals (300ms) and not only for brief signals (less than 10ms), as previously suggested by van den Brink and Houtgast (1989) [21]. In 1990, van den rink and Houtgast [24] made additional experiments on spectro-temporal integration, finding a trade-off between bandwidth and duration of the signal as follows: the integration is efficient for brief signals (even for wide spectrum signals) or for some longer signals (100ms) but only if the bandwidth is below the critical band. This suggests that both increasing the length of the signal and the bandwidth, deteriorate the efficiency of integration. They also investigated the effect of spectral gaps, finding that over a total bandwidth of up to 3 octaves, spectral gaps of 1/3 octaves do not deteriorate the efficiency of integration.

Nevertheless, the experiment reported here used relatively long signals (300ms) with wide spectrum (Approx. 5 critical bands) and the integration showed to be efficient. It is interesting to compare the experiments of van den Brink and Houtgast [21,24] with the experiment here described in order to assess the reasons for the different results. The main differences in experimental conditions were:

- [21,24] used pink and white noise as masker and not a pure tone.
- [21,24] used deterministic and not random targets.
- [21,24] used thresholds criteria defined for detection in 79.4 % and not 70.7 % of the trials. This affects the value of the constant k in a klog n rule.
- In [21,24] the total bandwidth of the compound stimulus grows when more bands are added. In the experiment reported here, the bandwidth was kept almost constant for the stimuli with 6, 11 and 21 bands. For the stimuli with 1 and 2 bands the bandwidth was much narrower. Nevertheless, the obtained slope was constant and hence the difference in the bandwidths of the stimuli with 1 and 2 bands against those with 6, 11 and 21 bands does not affect the 8.2log n result. It appears that the integration efficiency does not change in the range from 100 to 1000 Hz of BW.
- In [21,24] n is defined as the number of 1/3 octave bands presented simultaneously, while in this experiment n is the number of equal - bandwidth bands (or number of equal energy stimuli) presented simultaneously. Nevertheless, their definition of constant E/N_0 allows for direct comparison (at least in this respect).

The previous points suggest that the reason for the different results may dwell in the use of different signals. It could well be that the integration remains efficient over wider frequency range and longer time for non-deterministic targets than for deterministic ones. Green (1967) [26] supports the idea that the processing of a signal when masked by a noise and when masked by a sine wave are different. Theoretically, the use of a different threshold criterion affects the quantitative results but not the qualitative behavior.

9. Addition Effects in Masking

In order to predict whether or not a given complex signal is masked by another signal, several matters must be considered. First, the masking of a complex signal can be predicted from the masking of small parts of it only if the rules governing the addition of the masking of these small parts are known. Second, it is not possible to predict masking from the masker alone nor from the target alone. It is necessary to know the characteristics of both. Finally, from experiments on the addition of maskers [25] and the results presented here for addition of targets, it is clear that maskers as well as targets add, but they do so following different rules. Thus, in order to predict masking of a complex signal by another complex signal, the rules governing the addition of maskers and the addition of targets are necessary. Alternatively, masking could be predicted by using an excitation patterns approach [22] and a *decision rule*, but not knowing a good and general decision rule limits this approach.

In the previous chapters several experiments have been performed in order to asses the rules governing the additivity of targets. Nevertheless the additivity of maskers has bee ignored until now. Multiple experiments have been performed on the masker's additivity thus, it was not worth spending time experimenting in topics already studied. Humes and Jesteadt (1988) [25] presented a review of three models which are representative of the three main approaches for predicting masker's additivity.

The statistical characteristics of the maskers affect masking. Moore (1985) [27] studied the relation of the "excess" masking (extra masking produced by the addition of the maskers) and the statistical nature of the maskers (correlated or uncorrelated bands of noise). He found that uncorrelated maskers add more efficiently (3 to 9 dB) than correlated maskers (3 to 5 dB).

9.1 Additivity of Maskers

The first and simplest model is the linear energy-summation model. In this model, the threshold shift produced by combining two maskers, is assumed to be determined exclusively by the addition of the energy of the maskers. For example: if two maskers of equal energy are combined and the masking produced by one of them is x dB. Then, the combined target is expected to produce x + 3 dB of masking. If this rule were accurate, when two maskers are presented together the maximum addition expected would be 3 dB. Nevertheless, several experiments [10,25] have shown that more than 3 dB of masking addition may occur and the simple power-addition rule fails to account for this.

Failing to predict masking additivity with a linear addition of energy, the efforts moved towards more compressive rules that could explain more than 3 dB of masking addition (additions of the order of 10 dB have been documented [10,25]). The high-compression model and the power-law model (which can be considered as a particular case of the former) developed by Penner (1980) and Lutfi (1986) respectively, are useful and were shown to predict accurately addition of two or more non-simultaneous maskers and two or more simultaneous maskers, respectively. In 1988, Humes presented the modified power-law model with compressed internal noise. This model showed to be useful both for simultaneous and non-simultaneous masking and accounts properly for the effects found with low level maskers near to absolute threshold. Because of its relative simplicity and almost equivalent performance (the high compression model produces slightly better predictions) the modified power-law with compressed internal noise turns to be more suited for implementation in subband coding. A brief review of the later model follows. For more information on the topic and further details the reader is referred to Humes and Jesteadt (1988) [21].

Given two maskers A and B, each one of them individually producing an amount of masking (or masking shift with a tone as target) of MA and MB dB at f, (a particular frequency) respectively, when presented in isolation. With TH equal to the threshold in quiet. Then the amount of masking produced by A and B together (MA+B) can be calculated from:

$$[10^{(MA+B)/10}]^{0.3} = [10^{(MA+TH)/10}]^{0.3} + [10^{(MB+TH)/10}]^{0.3}$$

where:

$$[10^{(MA+TH)/10}]^{0.3} = [10^{MA/10}]^{0.3} + [10^{TH/10}]^{0.3}$$

and:

$$[10^{(MB+TH)/10}]^{0.3} = [10^{MB/10}]^{0.3} + [10^{TH/10}]^{0.3}$$

9.2 Additivity of Targets

The additivity of targets has been the main concern of the experiments here reported. From the results and discussions in the chapters 5, 6 and 7 the following points are summarized.

- When two targets are presented simultaneously, the threshold of the combined target is reduced. That is, the amount of masking is reduced.

- For two targets combined, the amount of the reduction in masking ranges between 1.5 dB and 3 dB. An important difference with respect to the addition of maskers is that when adding two targets, the reduction never is more than 3 dB, whereas two maskers may add more than 3 dB.

Furthermore, from the data collected it can be shown that the additivity of two targets of different level can be accounted for by a simple linear energy-summation rule. If TA and TB are the levels of targets A and B in dB SPL, the reduction in the threshold when both are presented simultaneously (with respect to the threshold of the strongest) is given by the following equations:

if TA is larger than TB:

$$\text{Reduction in dB} = K\,(10\log\,(10^{TA/10}+10^{TB/10}) - TA)$$

if TB is larger than TA:

$$\text{Reduction in dB} = K\,(10\log\,(10^{TA/10}+10^{TB/10}) - TB)$$

with K being a constant dependent on the characteristics of the targets. For most cases K is approx. 0.66 and has a range between 0.5 and 1. K=1 is equivalent to perfect integration (or having both targets in the same detection channel) while K=0.5 is equivalent to a 5log n addition rule with n=2. This model complements the discussion in section 7.3. For a larger number of targets, the experiment presented in chapter 8 and an additional experiment not described in this report, support the 8.2log n rule as valid for predicting the threshold shift produced by the presentation of n equal bandwidth bands in close to equally detectable conditions (as explained in chapter 8), even when the bandwidth of the complex target is not kept constant (in the range from 1 Hz to 1 kHz BW). A suggestion of how the rules for addition of maskers and targets could be used in subband coding follows in the next chapter.

10. Conclusion

10.1 The Use of Additivity Rules in Subband Coding

Figure 10.1 depicts a situation similar to that depicted in figure 4.2. A multiple-component stimulus is encoded and decoded. Again, each component produces masking in its surrounding frequency region as represented by the thick, solid lines. But now, the hatched bars represent a situation in which the quantization noise in each subband was allowed to grow as far as it was below the masked threshold produced by the signal. The picture was generated by empirically reducing the number of bits per sample for each subband as much as possible without degrading the subjective quality of the signal. In this case, the effect of masker's summation and target's summation is taken into account. As can be observed, the level of the quantization noise bands is consistently below the masking level for a single tone as target (thick lines). The bands of quantization noise (targets) effectively add as have been described before in this report. The spectrum of the just masked quantization noise is close to parallel to the masking pattern of the signal for a pure tone (then, the condition of close to equal detectability of the quantization noise is fulfilled). The bands have equal bandwidth. These conditions are similar to those in which the 8.2log n rule presented in chapter 8 was derived. Thus, applying the rule with n=5 should give a value close to the difference in levels observed in 10.1. Indeed, 8.2log 5 gives 5.73 dB which is close to the depicted in figure 10.1. It is not possible to assess the general correctness of the model with such a particular example, but the result is consistent with the empirical optimization given in the figure 10.1.

In order to apply the rules presented here in subband coding, two paths may be followed:

First, approximate the masking pattern of the signal by using the modified power-law and the individual masking patterns of the signal's components. The individual masking patterns of the signal's components may be calculated as described in chapter 6 or alternatively, as given in [10] or [22]. Then, given the number of bands of equal bandwidth used by the particular implementation of subband coding. An estimation of how many dBs below the signal's masking pattern each one of the bands of quantization noise may be allowed to go, is given by the 8.2log n rule. All the noise bands should be an equal number of dB below the signal's masking pattern in order to ensure consistency with the deduction of the rule (close to equal detectability as defined in chapter 8 and section 7.1.1).

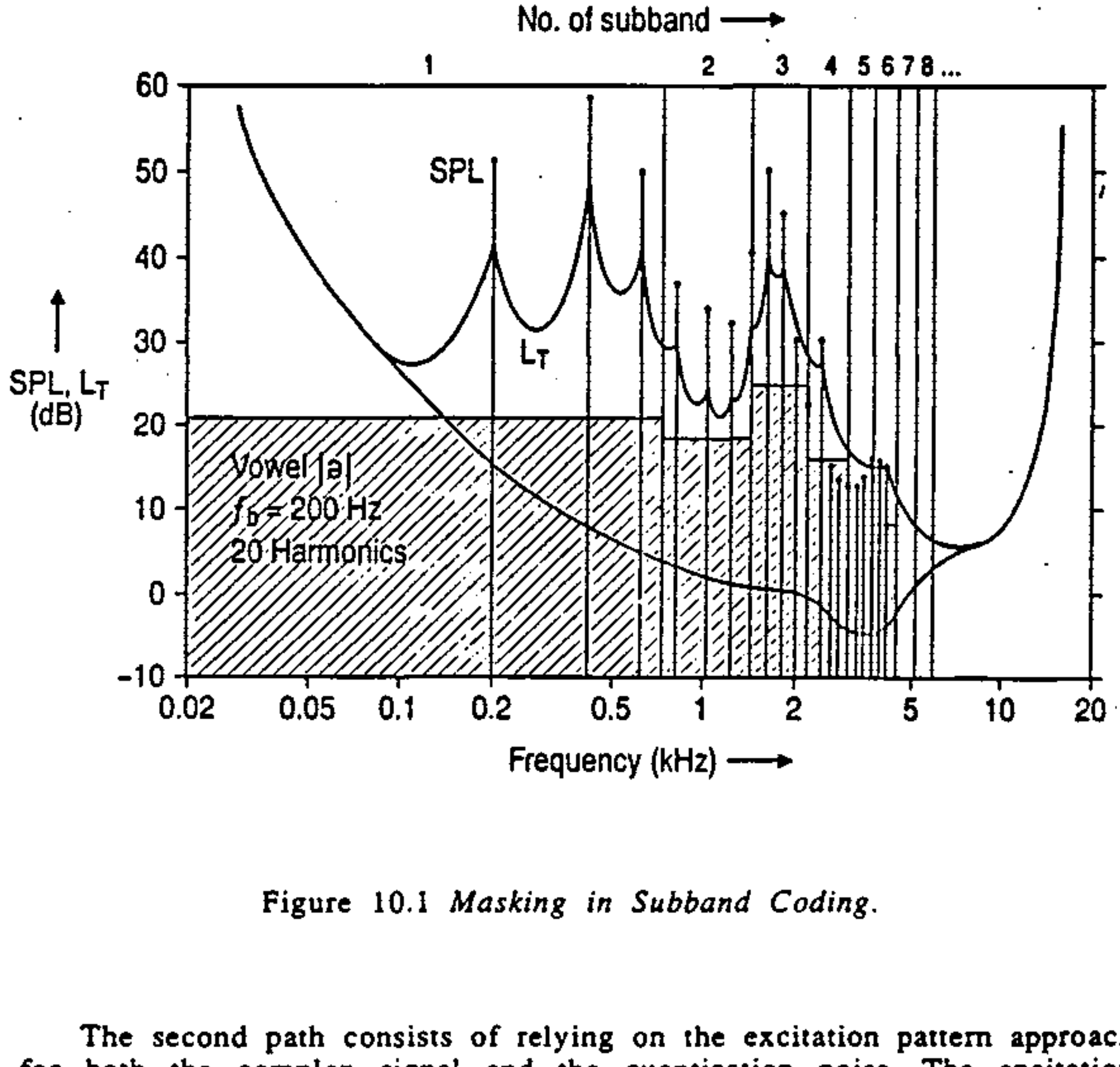


Figure 10.1 *Masking in Subband Coding.*

The second path consists of relying on the excitation pattern approach for both the complex signal and the quantization noise. The excitation pattern of the complex signal and the excitation pattern of the quantization noise can be calculated by using the rules summarized in chapter 9: modified power-law for the addition of the excitation patterns of the signal's components and simple linear energy-summation for the addition of the excitation patterns of the noise bands. The next step consists of the use of a *decision rule* to determine if the quantization noise is masked or not, and an optimization procedure to find the allocation which uses the minimum total number of bits, while keeping the noise masked. The well known problem of finding a good and general decision rule, and the necessity of a cumbersome optimization procedure, give bases to select the first path described.

M. del Río
May 1991

Appendix A

Timing of Trials and Signal Smoothing

Using a three-interval forced-choice procedure requires three observation intervals per trial (OI). In this particular implementation, each observation interval lasts 300ms and between each two OIs, a waiting time of 350ms was used. Additionally, before the first OI, a delay of 500ms was inserted for allowing the subject to assimilate the feedback from the previous trial and to prepare for the next trial. The process is cyclic and is best explained by figure A.1.

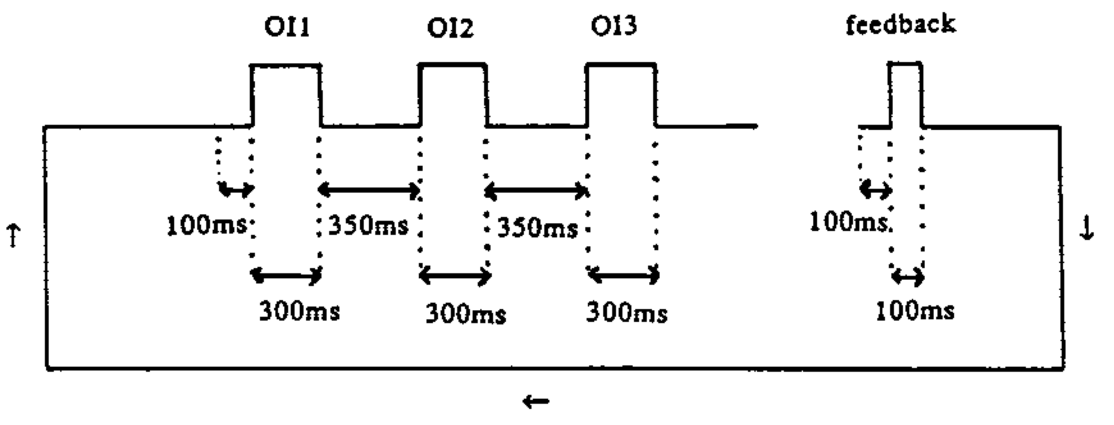


Figure A.1 *Trials and Timing.*

During each observation interval of 300ms, a signal containing the masker and sometimes the target (with a relative frequency of 1/3) was presented to the subject. In order to avoid spectral spread of the signal due to fast transitions at the beginning and end of each OI, a smoothing function was included in the attack and the decay of the stimulus presented in each OI. A rise cosine function was implemented with a time constant of 20ms as depicted in figure A.2. This smoothing function avoids bothersome "plops" for the subject, and spectral spread of the stimulus. was calculated to be better than 4 Hz, defining spectral resolution as the frequency interval between the first zero crossings in the spectrum on both sides of the tone. This resolution is more than enough for the experiments performed.

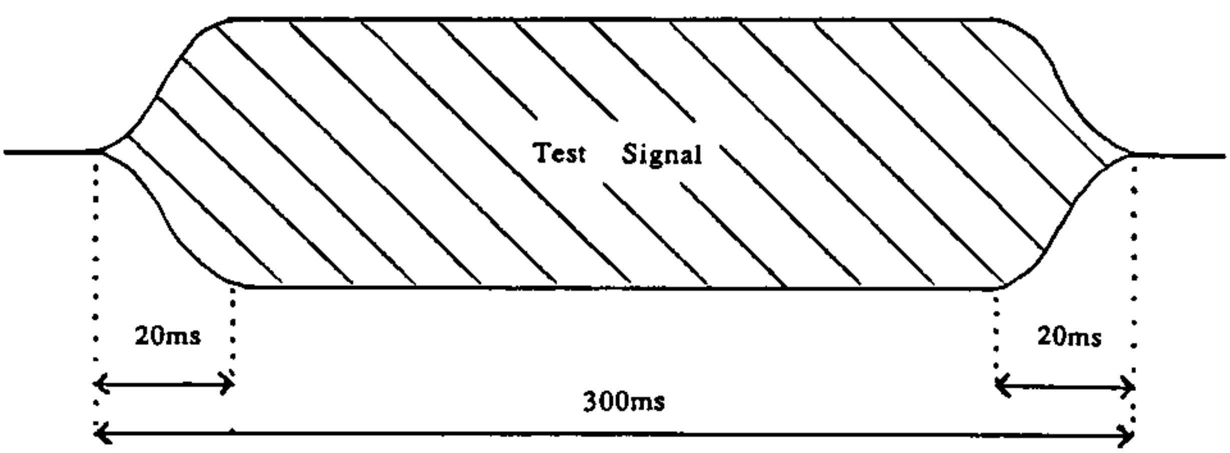


Figure A.2 *Duration, attack and decay of the Stimuli*

Appendix B

Combination Tones

When two tones are presented to the ear simultaneously, non linearities in the auditory system give rise to distortion products. These distortion products, even though not part of the acoustic stimulus, are present in the internal excitation and are perceived as part of the acoustic stimulus. Given two pure tones f_1 and f_2, the frequencies at which the distortion products or combination tones f(k) appear, as given by Scharf and Houtsma (1986) [4] are:

$$f(0) = f_2 - f_1$$

and

$$f(k) = f_1 - k(f_2 - f_1) \qquad \text{with k integer} \geq 1$$

Where f(0) is called the difference tone or quadratic distortion product and the tones f(k) with k ≥ 1 are the odd-order distortion products (cubic for k=1, fifth order for k=2, etc.). From Zwicker's (1981) and Goldstein's (1967) findings, f(0) behaves largely as a simple quadratic distortion component, whereas higher order combination tones exhibit frequency dependence. Zwicker (1981) proposed analytical equations to calculate the psychoacoustical excitation level of the different combination tones (CTs). His equations are based on measurements of the CTs by the cancellation method. If L(k) is the level of the combination tone at f(k) and L1 and L2 are the SPLs of the primaries f_1 and f_2 with $f_2 > f_1$ and $f_1 > 200$ Hz.

$$L(0) = L1 + L2 - 126 + 10\log[1+(f_1^2/16)] + 10\log[1+(1/4f_1^2)]$$

Where L(0) is the level of the difference tone or quadratic distortion product. For a given combination of primary frequencies, L(0) increases monotonically when the level of one or both of the primaries is increased. For L(1) the excitation level is given by:

$$\begin{aligned} L(1) = 1.1L1 - 10\{ \; & \log[1+10^{0.2(L1-L2)+0.6(0.1L2-\Delta Z^2-7)}] \\ & + \log[1+10^{0.05(L2-L1)(1-0.35\Delta Z)}] \\ & + \log[1+10^{0.04(L2-40)}] \\ & + \log[1+(f_1^2/16)] + \log[1+(1/(4f_1^2)] \\ & + 1.2\ \Delta Z^2[1-0.01(L2-50)] \; \} - 2 \end{aligned}$$

Where ΔZ is the difference in barks between f_1 and f_2, which can be approximated by:

$$\Delta Z = 14.2\log(f_2/f_1)$$

The equation for L(1) is valid for ratios of f_2/f_1 smaller than 1.3 and was obtained from data with f_1 larger than 1.6 kHz. Similar approximations for the levels L(2) and L(3) are given by:

$$L(2) = L(1) - \Delta Z - 18$$

$$L(3) = L(1) - 2\Delta Z - 30$$

The effects of the combination tones have been observed in the neural correlates at the eigth nerve and superior processing levels, suggesting its origin in a previous processing stage. The odd-order combination tones, f(1) and higher, have been attributed to cochlear mechanics while the difference tone $f_2 - f_1$ seems to be produced at the middle ear level before the frequency selectivity of the auditory system is manifested.

When a band of noise is presented to the ear, multiple combination tones are generated at frequencies given by all the possible combinations of the components of the band. From these combination tones, those produced at $2f_1$-f_1 are the most intense. The ensemble of CTs effectively extend the excitation pattern of the noise band towards the low frequencies, as given by Greenwood (1971) [16]. In general, a complex signal generates a complex ensemble of CTs which become present in the internal excitation pattern and influence the perception of the signal. These ensembles of CTs which aggregate to the complex signals are called combination aggregates [16].

Appendix C

Software Utilities

The next is a brief list of the software available at the Institute for Perception Research, The Netherlands, which was extremely useful for the fulfillment of this project.

- DISSPLA Used to generate the plotts of the masking patterns from the experimental results.

- WordPerfect Used to create this report.

- Software for controlling the IEEE 488 interface and equipment.

- Software to control and perform files transference to the DSC 200.

The following software was developed by the author of this report with the advisory of A.J.M. Houtsma and Th. de Jong, also at the Institute for Perception Research.

To estimate thresholds of pure tones and noise bands of 50, 100 and 200 Hz in quiet or in noise:

THRES
THRES50
THRES100
THRES200

To generate bands of noise with arbitrary bandwidth, cut-off frequencies, sampling rate and duration and "constant" energy:

NOISEGEN

To estimate masked thresholds of tones or noise bands of 50, 100 and 200 Hz or combinations of noise bands using a tone as masker and a three-interval forced-choice procedure:

STV3 (single tone)
SBNV3 (single band of noise)
TBNV3 (two bands of noise)
FULL (arbitrary number of noise bands with arbitrary inter-band level differences)

For calculation of auditory excitation patterns in accordance to Glasberg and Moore (1990) [22]:

XPATT89

References

[1] Green, D.M. (1976), "An introduction to hearing", Lawrence Erlbaum Associates, Publishers, distributed by John Wiley & Sons, New York.

[2] Rossing, T.D. (1982), "The Science of Sound", Addison-Wesley, Reading, MA.

[3] Scharf, B. and Buus, S. (1986), "Audition I: Stimulus, Physiology, Thresholds" in Handbook of Perception and Human Performance, ed.K.R. Boff, L. Kaufman and J.P. Thomas, J. Wiley, New York. pp 14.1-71.

[4] Scharf, B. and Houtsma, A.J.M., (1986), "Audition II: Loudness, Pitch, Localization, Aural Distortion, Pathology" in "Handbook of Perception and Human Performance", ed. K.R. Boff, L. Kaufman and J.P. Thomas, J. Wiley, New York. pp 15.1-60.

[5] Zwislocki, J. (1965), "Analysis of Some Auditory Characteristics" in "Handbook of Mathematical Psychology", volume III, ed. R.D. Luce, R.R. Bush and E. Galanter, John Wiley & Sons, New York. pp 1-97.

[6] Verschuure, J. (1978), "Auditory Excitation Patterns", doctoral thesis, University of Delft, The Netherlands.

[7] Kim, D.O. and Parham, K. (1991), "Auditory nerve spatial encoding of high-frequency pure tones", Hearing Research, vol. 52, 167-180.

[8] Kiang, N.Y. (1965), "Discharge of single fibers in the cat's auditory nerve", Research Monograph No. 35, Cambridge, Mass., MIT Press.

[9] Levitt, H. (1971), "Transformed Up-Down Methods in Psychoacoustics", JASA, vol. 49, 467-477.

[10]Zwicker, E. and Fastl, H. (1990), "Psychoacoustics", Springer-Verlag, Berlin.

[11]Schlauch, R.S. and Rose, R.M. (1990), "Two-, three-, and four-interval forced-choice staircase procedures: Estimator bias and efficiency", JASA, vol. 88, 732-740.

[12]Veldhuis, R. (1991), "Backgrounds of audio source coding", unpublished, Philips Research Laboratories, Eindhoven, The Netherlands.

[13]Subramaniam, M. *et al.* (1990),"Evoked-Response Tone-on-Tone Masking in the Chinchilla: Effect of Masker Frequency", Audiology, vol. 29, 202-211.

[14]Ehmer, R.H. (1959), "Masking by Tones *vs* Noise Bands", JASA, vol. 31, 1253-1256.

[15]Goldstein J.L. (1967), "Auditory nonlinearity", JASA, vol. 41, 676-689.
[16]Greenwood D.D. (1971), "Aural Combination Tones and Auditory Masking", JASA, vol. 50, 502-543.
[17]Smoorenburg, G.F. (1972), "Combination Tones and Their Origin", JASA, vol. 52, 615-632.
[18]Zwicker, E. (1980), "Reversed behavior of masking at some levels", Audiology, vol. 19, 330-334.
[19]Wegel, R.L. and Lane, C.E. (1924), "The auditory masking of onepure tone by another and its probable relation to the dynamics of the inner ear", Physical Rev., vol. 23, 266-285.
[20]Isabelle, S.K. and Colburn, H.S. (1991), "Detection of tones in reproducible narrow-band noise", JASA, vol. 89, 352-360.
[21]van den Brink, W.A.C. and Houtgast, T. (1990), "Efficient across-frequency integration in short-signal detection", JASA, vol. 87, 284-291.
[22]Glasberg, B.R. and Moore, C.J. (1990), "Derivation of auditory filter shapes from notched-noise data", Hearing Research, vol. 47, 103-138.
[23]Egan, J.P. and Hake, H.W. (1950), "On the Masking Pattern of a Simple Auditory Stimulus", JASA, vol. 22, 622-630.
[24]van den Brink, W.A.C. and Houtgast, T. (1990) "Spectro-temporal integration in signal detection", JASA, vol. 88, 1703-1711.
[25]Humes, L. and Jesteadt, W. (1989), "Models of the additivity of Masking", JASA, vol. 85, 1285-1294.
[26]Green, D.M. (1967), "Additivity of Masking", JASA, vol. 41, 1517-1525.
[27]Moore, B.C.J. (1985), "Additivity of simultaneous masking, revisited", JASA, vol. 78, 488-494.