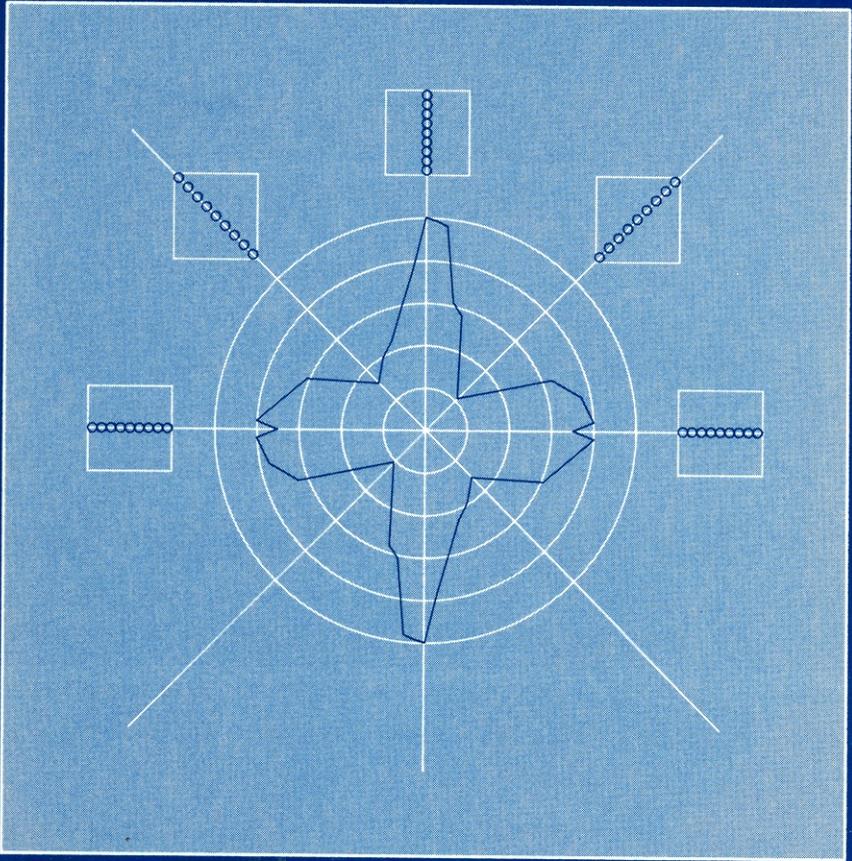


ASPECTS OF SPECTRO-TEMPORAL PROCESSING IN AUDITORY SIGNAL DETECTION



Willem A.C. van den Brink

**Aspects of spectro-temporal processing
in auditory signal detection**

**Aspecten van spectro-temporele verwerking
bij auditieve signaaldetectie**

(met een samenvatting in het Nederlands)

proefschrift

ter verkrijging van de graad van doctor aan
de Rijksuniversiteit Utrecht, op gezag van
de rector magnificus prof.dr. J.A. van Ginkel,
volgens besluit van het college van decanen
in het openbaar te verdedigen op donderdag
16 mei 1991 des namiddags te 14.30 uur.

door

Wilhelmus Albertus Christianus van den Brink

geboren op 20 maart 1958 te Gendt

Bariet, Ruinen

1991

Promotor : prof.dr. G.F. Smoorenburg

Co-promotor : dr.ir. T. Houtgast

The work presented in this thesis was performed in the Department of Audiology, TNO Institute for Perception in Soesterberg.

This research has been financially supported by the Netherlands Organization for Scientific Research (NWO).

VOORWOORD

Het in dit proefschrift beschreven onderzoek werd uitgevoerd, in de periode van 15 oktober 1986 tot 15 oktober 1990, bij de afdeling Audiologie van het TNO-Instituut voor ZintuigFysiologie te Soesterberg, in tijdelijke dienst van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek.

Gaarne wil ik iedereen die heeft bijgedragen aan de totstandkoming van dit proefschrift bedanken.

De collega's op de afdeling Audiologie, Evert Agterhuis, Els van Amerongen, Wim Bles, Adelbert Bronkhorst, Frank Geurtsen, Bernd de Graaf, Tim Marcus, Ton Mimpfen, Reinier Plomp, Louis Pols, Hans van Raaij, Herman Steeneken, Jeroen van Velden, Jan Verhave en Joos Vos wil ik bedanken voor hun ondersteuning en de prettige werksfeer.

De leden van de promotiecommissie prof.dr. A.J.M. Houtsma, prof.dr. J.J. Koenderink, prof.dr. D. van Norren, prof. dr. ir. R. Plomp en prof.dr.ir. M.A. Viergever wil ik bedanken voor hun bereidheid het proefschrift te beoordelen.

Mijn promotor Guido Smoorenburg wil ik bedanken voor zijn vele adviezen en de stimulerende discussies die mij steeds weer een andere kijk op de materie gaven.

Mijn co-promotor Tammo Houtgast wil ik in het bijzonder bedanken, als directe begeleider van het onderzoek was hij voor mij een constante stimulans, met name de maandagavond meetsessies voorafgegaan door een warme maaltijd met zijn gezin zullen mij altijd bij blijven.

CONTENTS

INTRODUCTION

CHAPTER 1 EFFICIENT ACROSS-FREQUENCY INTEGRATION IN SHORT-SIGNAL DETECTION

- 1.1 Introduction
- 1.2 Stimuli and procedure
 - 1.2.1 Signal Definition
 - 1.2.2 Apparatus
 - 1.2.3 Level definition
 - 1.2.4 Subjects
 - 1.2.5 Procedure
- 1.3 Experiments with narrowband signals
- 1.4 Experiments with broadband signals
 - 1.4.1 Equal- $(E/N_0)_{1/3}$ signals
 - 1.4.2 Dirac pulses in white noise
 - 1.4.3 Control experiment
 - 1.4.4 Noncontiguous signal spectra
 - 1.4.5 Stimulus Level
- 1.5 Discussion
- 1.6 Conclusions

CHAPTER 2 SPECTRO-TEMPORAL INTEGRATION IN SIGNAL DETECTION

- 2.1 Introduction
- 2.2 Stimuli and procedure
 - 2.2.1 Signal definition
 - 2.2.2 Level definition
 - 2.2.3 Apparatus
 - 2.2.4 Subjects
 - 2.2.5 Procedure
- 2.3 Experiments
 - 2.3.1 Two durations and two bandwidths
 - 2.3.2 Restricted spectro-temporal surface
 - 2.3.3 Critical bandwidth for efficient temporal integration
 - 2.3.4 Critical time window for spectral integration
- 2.4 General discussion
- 2.5 Summary and conclusions

CHAPTER 3 SIGNAL DETECTION IN TEMPORALLY MODULATED AND SPECTRALLY SHAPED MASKERS

- 3.1 Introduction**
- 3.2 Methods**
 - 3.2.1 Procedure**
 - 3.2.2 Subjects**
 - 3.2.3 Apparatus**
 - 3.2.4 Level definition**
- 3.3 Noise bandwidth**
 - 3.3.1 Stimuli**
 - 3.3.2 Results and discussion**
- 3.4 Across-frequency integration of signal energy and CMR**
 - 3.4.1 Control experiment**
 - 3.4.2 Short-duration signals**
 - 3.4.3 Long-duration signals**
 - 3.4.4 Un-comodulated masking condition (UN-COMOD)**
- 3.5 Temporal resolution and the width of the ear's temporal window**
 - 3.5.1 Stimuli**
 - 3.5.2 Results and discussion**
- 3.6 The spectral analog of CMR**
 - 3.6.1 Masker duration**
 - 3.6.2 Coshaped versus un-coshaped masking condition**
- 3.7 Spectral resolution and the auditory bandwidth**
 - 3.7.1 Stimuli**
 - 3.7.2 Results and discussion**
- 3.8 General discussion**
- 3.9 Conclusions**
- Appendix**

SUMMARY

SAMENVATTING

REFERENCES

CURRICULUM VITAE

INTRODUCTION

This thesis presents a psychophysical study concerned with human auditory signal detection in noise. It is focussed on the ear's ability to integrate stimulus energy distributed across frequency and across time. Below the general background and rationale for this study will be briefly discussed.

In every day life a wide variety of sounds are perceived by the human auditory system. Considering the complexity of the incoming sounds it is surprising to see how well the human observer copes with all this information. One aspect of sound perception that was of particular interest to us was the fact that we can very well discriminate sounds in noisy environments. How does the human auditory system manipulate the incoming sounds in order to detect signals in a noisy environment? This question can be studied under laboratory conditions, using sounds that are well defined. In these so-called psycho-acoustical experiments input-output relationships based on responses of subjects to sound stimuli are obtained. Together with physiological data, a functional description of the human hearing system can be obtained.

Currently, the human peripheral auditory system is modelled as containing an array of bandpass filters. These auditory filters appear to play an important role in many aspects of auditory perception, including masking, loudness perception, timbre perception and pitch perception (for reviews, see Plomp, 1976, and Moore, 1989). Given the frequency selectivity of the ear, the incoming sound is often characterized by its spectrogram; a pattern that depicts the frequency spectrum of the stimulus, averaged over short-time intervals, as a function of time. Typically, the degree of frequency selectivity is restricted to 1/3-octave bands, in line with the auditory "critical bandwidth" as manifest in a large body of psychoacoustical data.

In the past, much attention has been paid to signal detection experiments concerned with temporal integration. Detection thresholds of pure-tone bursts were determined against a background of Gaussian noise, as a function of tone-burst duration. In general detection thresholds are, up to tone burst durations of about 150-200 ms, determined by the tone-burst energy; the tone-burst intensity is efficiently integrated over time. These experiments typically reveal auditory processing within one critical band: *within-channel* processing. However, natural sounds like speech show, as illustrated by their spectrogram, a distribution of sound intensity across both time and frequency.

In general, a sound stimulus may activate various auditory channels simultaneously. Thus, besides within-channel processing for narrow-band signals the perception of broad-band signals involves some kind of *across-channel* processing. The experiments reported in this thesis present a systematic study on across-channel processing in auditory signal detection. In a typical experiment we determined the detection thresholds of various spectro-temporal signal patterns presented in continuous broadband noise. These signal patterns have been systematically varied in order to study, within one experimental paradigm, both *within-channel* and *across-channel* processing.

In case of broad-band signals, classical signal detection experiments suggested that the auditory channel with the best signal-to-noise ratio determines the detection threshold (e.g. Fletcher, 1940, Gässler, 1954). This model has often been referred to as the power-spectrum model. Although this model could explain a number of experimental results, other experiments (e.g. Green, 1958, Scholl, 1962) already showed that signal energy distributed across auditory channels was integrated in the detection process. Scholl (1962) used bandpass-filtered noise bursts and found that duration is a critical parameter for the effect of bandwidth on the detection threshold. The experiments as presented in the first chapter of this thesis are concerned with across-channel integration of signal components simultaneously presented to different auditory channels. It will be shown that for short-duration signals the across-channel, or across-frequency, integration is almost as efficient as the classical within-channel temporal integration mentioned above. The second chapter will show, in more detail, for what stimulus conditions, in terms of bandwidth and duration, this type of efficient processing is observed.

The relation between across-frequency integration as mentioned above and other across-channel processing phenomena are the subject of the third chapter. In a series of experiments performed by Hall and colleagues (Hall et al 1984; Hall 1986) an interesting across-channel processing phenomenon in auditory signal detection was shown; the masking of a pure-tone signal presented in a modulated narrow-band noise masker decreased when a second modulated narrow-band masker was added remote in frequency from the first masking band. This release from masking occurred only when the two noise bands had similar temporal envelopes and was thus called co-modulation masking release (CMR). The current models accounting for CMR are based on the assumption that the across-channel processing phenomenon results from across-channel envelope disparity. Experiments with respect to signal detection in broad-band temporally modulated noise are presented. The key question of

the third chapter of this thesis, is the interaction between on the one hand, the reduction in across-channel envelope disparity with increasing signal bandwidth which predicts a higher detection threshold, and, on the other hand, the effective across-frequency integration, reported in the first chapter, which would predict no such effect.

In general terms, the main issue concerns the effect of a signal's spectro-temporal pattern on its detection. The auditory processing characteristics in the frequency-time domain as revealed by this study are, of course, of interest for understanding audition in general: Many natural sounds show a great variety of spectro-temporal patterns, and the perception of these sounds is governed to some degree by the processing characteristics studied in this thesis.

Fletcher, H. (1940). "Auditory patterns," *Rev. Mod. Phys.* **12**, 47-61.

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

Green, D.M. (1958). "Detection of multiple component signals in noise," *J. Acoust. Soc. Am.* **30**, 904-911.

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

Hall, J.W. (1986). "The effect of across-frequency differences in masking level on spectro-temporal pattern analysis," *J. Acoust. Soc. Am.* **79**, 781-787.

Moore, B.C.J. (1989). "*An Introduction to the Psychology of Hearing, Third Edition*," Academic Press, London.

Plomp, R. (1976). "*Aspects of Tone Sensation*," Academic Press, London, 1-167.

Scholl, H. (1962). "Über die Bildung der Hörschwellen und Mithörschwellen von impulsen," *Acustica* **12**, 91-101.

CHAPTER 1

EFFICIENT ACROSS-FREQUENCY INTEGRATION IN SHORT-SIGNAL DETECTION

Willem A.C. van den Brink and Tammo Houtgast
Journal of the Acoustical Society of America 87, 284-291 (1990).

ABSTRACT

A series of experiments was performed on the influence of bandwidth on the masked threshold of brief deterministic signals in continuous broad-band noise. The signal bandwidth is quantified by the number (n) of constituent 1/3-octave bands. For n increasing from one to typically nine, the masked-threshold level in the constituent 1/3-octave bands is found to decrease by $8\log(n)$. This integration rule is obtained when each of the 1/3-octave bands covered by the signal equally contributes to detection, i.e. that for each of these 1/3-octave bands the difference between signal level and the individual masked-threshold level is the same. It was found that this integration rule also applies to noncontiguous signal spectra and that it remains intact over a broad range of masker levels. Commonly, the masked threshold of compound signals (for instance, n frequency components with a spacing of typically 1/3-octave), relative to the masked threshold of single component signals has been described by a $5\log(n)$ integration rule. However, this rule was obtained for signal durations of typically 100 ms or more. For the present brief signals (typically 10 ms or less) the across-frequency integration is found to be more effective.

1.1 INTRODUCTION

Recently, various phenomena related to across-frequency processing have been reported. Green et al. (1984) performed a series of experiments on 'profile analysis', using a paradigm which ensures that the data reflect global, across-frequency band processing rather than processing based on a single critical-band output. Also the phenomenon of 'Comodulation Masking Release' (CMR, Hall, 1985), indicating that synchronous fluctuations in frequency bands flanking the signal frequency facilitate signal detection, must involve some kind of across-frequency processing. Other examples involving across-frequency processing are 'Comodulation Difference Detection' (CDD) (McFadden, 1987) and 'Temporal-gap detection over frequency' (Grose and Hall, 1988).

In classical signal detection (i.e., the masked threshold of a signal in continuous broad-band noise), there still remain some questions concerning the possible role of across-frequency processing in case of broad-band signals. Classical results found by von Gässler (1954) indicated that within a critical band energy is integrated up to 100 - 200 ms and that the 'best' critical band essentially determines the masked threshold. Green (1958) however, proposed a statistical summation model to determine the masked threshold of a multi-component signal with respect to the threshold of each of the single-component stimuli. Essentially, this model yields a $5\log(n)$ integration rule, in which n specifies the number of signal components (usually sinusoidal components with a spacing of 1/3-octave band or more). According to this rule, within each excited 1/3-octave band, the masked threshold of a two-component signal is $5\log(2) = 1.5$ dB lower than that of a single-component signal, which was in agreement with his experimental data. Buus et al. (1986) compared the masked threshold of a 450 ms 18-tone complex relative to that of a single pure tone. They found that the difference in masked threshold followed the $5\log(n)$ integration rule. In earlier experiments performed by Scholl (1962) however, the significance of signal duration for bandwidth integration has been shown. Scholl used band-pass filtered noise pulses of various durations and found that duration is a critical parameter for the effect of bandwidth on the masked threshold. For very brief signals, the data suggest efficient across-frequency integration, i.e. a 3 dB decrease for each doubling of bandwidth which amounts to a $10\log(n)$ integration rule. Is this efficient across-frequency integration a consequence of the stochastic nature of the signals used (temporally gated and band-pass filtered noise), or is it also obtained for very brief deterministic signals?

The present experiments are concerned with the masked threshold of

brief deterministic signals, covering one or more 1/3 octaves. In order to fully mobilize the possible benefits of across-frequency integration in case of the broad-band signals, all the frequency regions covered by the signal should equally contribute to signal detectability.

1.2. STIMULI AND PROCEDURE

1.2.1 Signal definition

Three types of signals were used in the present experiments, which will be referred to as Single Gaussian, Compound Gaussian and Complex Signals, respectively.

1. Single Gaussian.

These signals consist of a tone pulse with a Gaussian envelope, being defined by:

$$s(t) = C \sin(2\pi f_0(t-t_0)) \exp(-\pi(\alpha f_0(t-t_0))^2) \quad (1)$$

The value of f_0 determines the center frequency of the tone pulse. The value of α determines the effective duration ($t_{\text{eff}} = 1/(\alpha f_0)$) and the effective bandwidth ($b_{\text{eff}} = \alpha f_0$) of the stimulus. We used $\alpha = 0.2$, for which the bandwidth is just within a 1/3-octave. The -6 dB duration for a single Gaussian signal for $f_0 = 500$ Hz is 9.5 ms. This duration reduces by a factor of 0.5 for each doubling of f_0 . The waveform of a Single Gaussian tone pulse and its spectrum are plotted in Fig. 1.

2. Compound Gaussian.

The compound Gaussian signals are obtained by adding single Gaussian signals with different center frequencies, typically at 1/3-octave intervals, with the peaks of all Gaussian envelopes coinciding in time. Again the value of α was 0.2 for each single Gaussian pulse. Thus by adding single Gaussian pulses of different center frequencies the spectral contents of the compound Gaussian signal was essentially equal to the sum of the spectra of the single Gaussian pulses, as illustrated in Fig. 2.

3. Complex Signals.

These signals are defined in the frequency domain at frequency intervals of typically 5 Hz, and the frequency components were added in cosine-phase. The waveform of a complex signal and its spectrum are plotted in Fig. 3. The

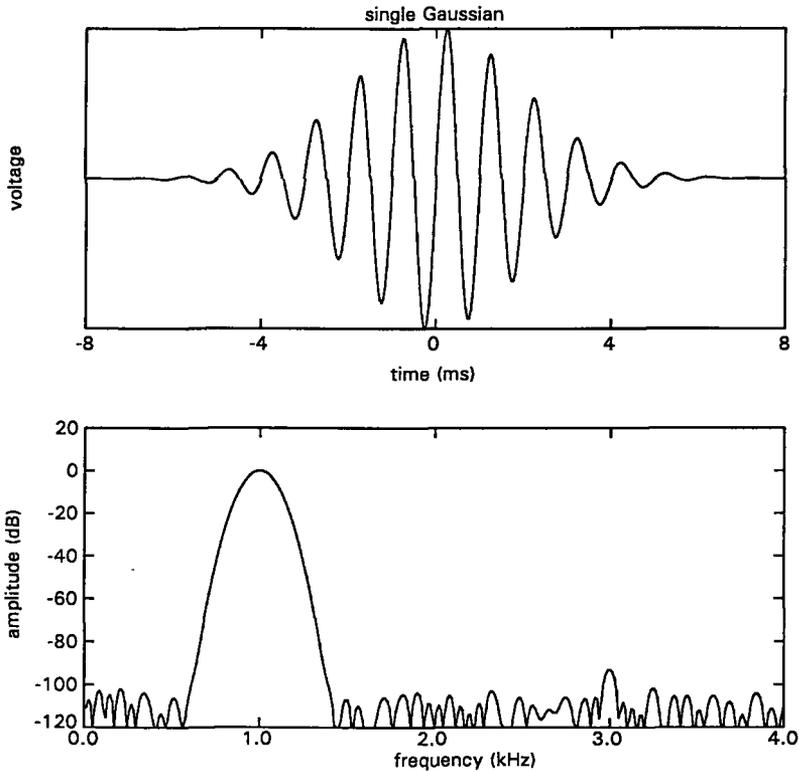


Fig.1. Waveform (upper graph) and spectrum (lower graph) of a single Gaussian signal, as defined by eq(1). For $f_0 = 1.0$ kHz the -6 dB duration for this signal is 4.8 ms.

effective duration depends on the bandwidth. For the smallest bandwidth considered (1/3 octave around 1.6 kHz) the -6 dB duration is about 3.2 ms.

1.2.2 Apparatus

The experiments were under control of a PDP 11/34 computer. Signals were generated with a digital signal processor based on a TMS32010, at 40 microsecond sample intervals, and with 10 kHz low-pass filtering. Broad-band noise was generated, using an analog Wandel & Golterman RG-1 noise generator, and was digitally mixed with the signals. The stimuli were presented diotically (same stimulus to both ears) by means of Beyer DT-48 headphones. Subjects were seated in a soundproof room.

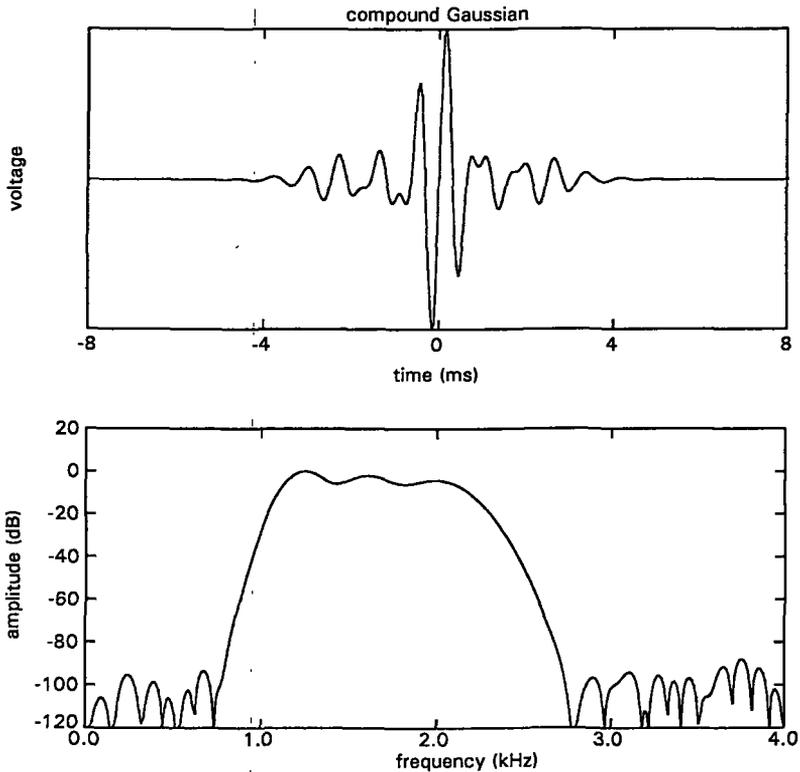


Fig.2. Waveform (upper graph) and spectrum (lower graph) of a compound Gaussian, composed of single Gaussian signals at 1.25, 1.6 and 2.0 kHz.

1.2.3 Level definition

The levels of the signals were defined by the energy within 1/3-octave bands ($E_{1/3}$). For the single Gaussian signals this corresponds to their actual energy (E). For broad-band signals the energy spectrum was determined after filtering with a 1/3-octave filter bank (Brüel & Kjaer 2112). The energy within a 1/3-octave band is always related to the spectral density of the masking noise within the same 1/3-octave band : $(E/N_0)_{1/3}$. N_0 is the mean noise energy within a time-frequency window of which the product of bandwidth and duration equals one or also the mean noise intensity in a 1-Hz interval.

Two types of masking noise were used: white noise (N_0 constant over f) and pink noise (N_0 decreasing in proportion to $1/f$). The level of the masking noise was always set at approximately 50 dB SL in all experiments apart from the experiment in section 1.4.5 : Stimulus Level.

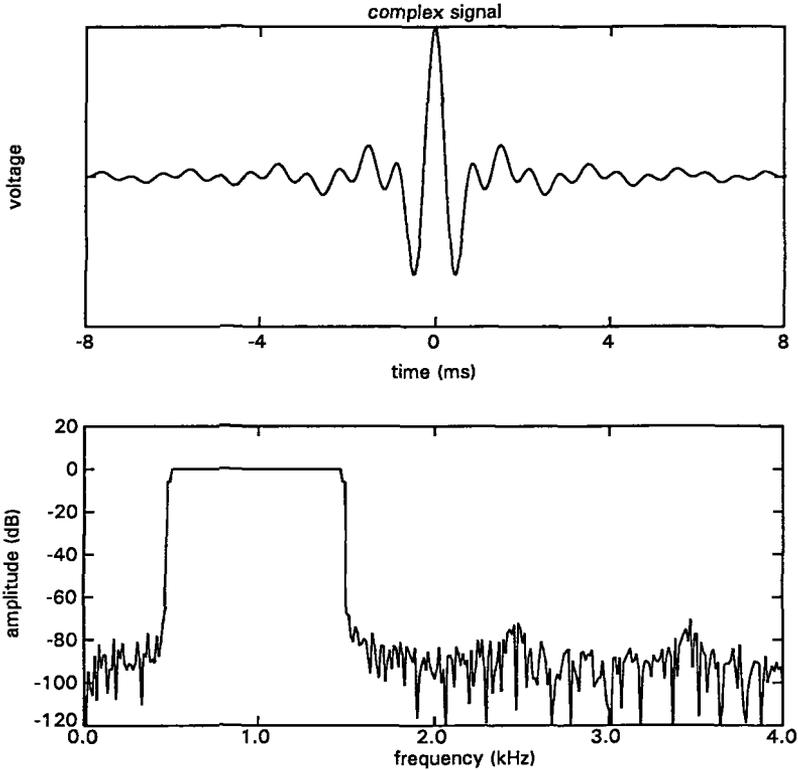


Fig.3. Waveform (upper graph) and spectrum (lower graph) of a complex signal. The -6 dB duration for this signal is about 1.7 ms.

1.2.4 Subjects

Four subjects participated in the experiments. Two were experienced in psychoacoustic experiments, the other two were students who were paid an hourly wage. The students were trained half a day before data collection started.

1.2.5 Procedure

Two types of 2-IFC procedures were used. In both procedures, each trial consisted of two 300-ms observation intervals, separated by 200-ms, and the signal was presented with equal probability in the first or second interval. Intervals were marked by a light at the listener's response box. Feedback was provided after each response. Broad-band noise was presented continuously. The first procedure was applied in order to obtain psychometric functions. In a series of trials signal presentation was randomized, i.e. for each trial a signal was selected at random from the set of signals considered in one series (for

instance with different center frequency as in section 1.3). After each trial the data for that signal were stored and a 'new' signal was chosen at random. So, separate adaptive tracks for the individual signals were interleaved, which was done in order to balance any possible effect of fatigue on the measured thresholds of the signals. The signal level was varied adaptively. Initially, the signal level was set well above masked threshold, and it decreased by 4 dB after each correct response. After the first incorrect response the level was increased by 4 dB and the stepsize was decreased to 2 dB: subsequently a 2 dB level decrease was applied after two successive correct responses and a 2 dB increase after each incorrect response. Within one series the number of trials for each condition was 30. Responses were stored for each signal level. The data from ten series were pooled and the psychometric functions, the percentage of correct responses as a function of signal level, were plotted (percentages were plotted only for those signal levels where at least 40 trials were obtained). The masked threshold level was obtained by interpolation at 75% score.

In the second (faster) method a three-down one-up procedure (Levitt, 1971) was applied in order to estimate the level which would produce 79.4% correct responses. Again for each trial a signal was selected at random and signal level adaptation was performed as in the first procedure described above. After three reversals data collection started up to a total of thirteen reversals. Threshold was defined as the mean of the signal levels at the last ten reversals. Typically, ten such threshold estimates were obtained and averaged. This second method was used in the experiments reported in Secs. 1.4.1-4.

1.3 EXPERIMENTS WITH NARROW-BAND SIGNALS

This experiment with only 1/3-octave-wide signals served as a baseline for the subsequent experiments with broad-band signals. Single Gaussian tone pulses were used as a signal (Sec. 1.2.1). The masked threshold was measured as a function of the central frequency of the pulse. For one frequency (1.0 kHz) the psychometric functions obtained for the four subjects are presented in Fig. 4. For comparison, these data are related to the classical results from detection experiments by Green (1957). Although for the ideal signal detector d' amounts to $d' = (2E/N_0)^{1/2}$, Green found that the experimental data were best described by $d' = k (E/N_0)$, with typically $k = 0.1$. The psychometric function corresponding to Green's description is included in Fig. 4. With respect to both the slope and the position of the function, the present data are

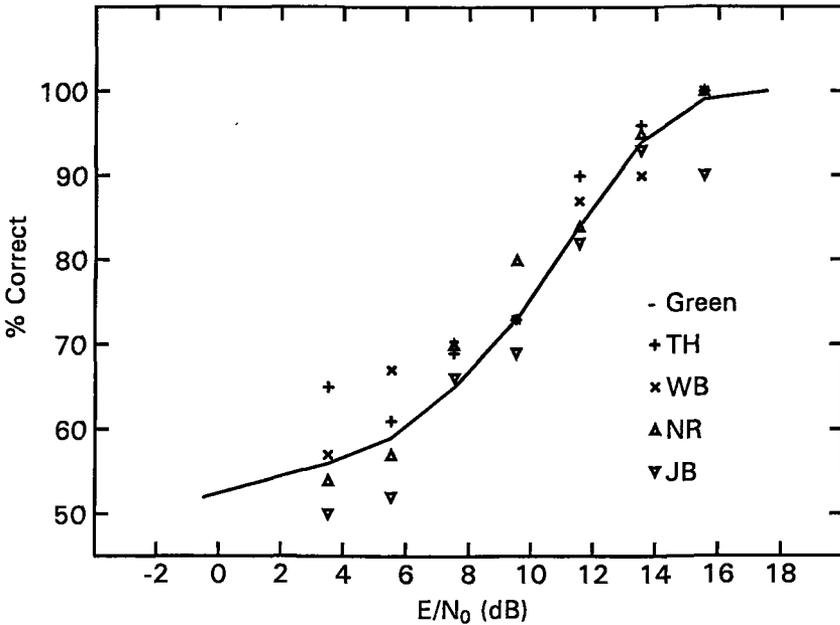


Fig.4. Psychometric datapoints for four subjects for a single Gaussian signal centered at 1.0 kHz. The curve represents the psychometric function according to Green (1957).

consistent with Green's data.

Fig. 5. presents the thresholds obtained for the different center frequencies. There is a tendency for the threshold to increase towards higher frequencies, yielding a difference in masked threshold of approximately 1 dB/octave. This is somewhat less than the slope of approximately 1.9 dB/octave as found by Green et al.(1959) for 100 ms pure tones. Still, as a first approximation E/N_0 may serve as a simple first order measure for detectability of brief narrow-band signals over a wide frequency range. Thus, the broad-band signals used in the next experiments were designed to meet the condition of equal E/N_0 for each one-third octave band covered by the signal : equal- $(E/N_0)_{1/3}$ signals.

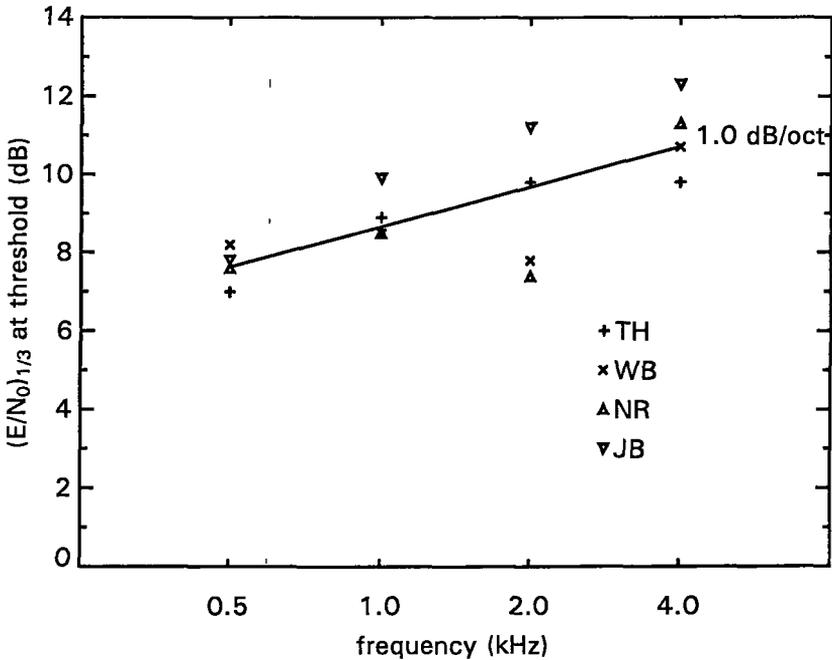


Fig.5. Masked threshold as a function of frequency for single Gaussian signals. The line represents a linear regression fit through the datapoints. The datapoints were extracted from psychometric functions.

1.4 EXPERIMENTS WITH BROAD-BAND SIGNALS

1.4.1 Equal- $(E/N_0)_{1/3}$ signals

With 1.6 kHz as the geometrical center frequency, a set of brief complex signals was designed using bandwidths of 1/3, 3/3, 5/3, 9/3 and 13/3 octaves. For sake of generality, two types of masking noise were used: white noise and pink noise. For white noise, the condition of equal $(E/N_0)_{1/3}$ found in the previous section is fulfilled when $E_{1/3}$ (the signal energy within a 1/3-octave band) is equal for all 1/3-octave bands covered by the signal. For pink noise, the condition of equal- $(E/N_0)_{1/3}$ requires $E_{1/3}$ to decrease by 1 dB for each next higher 1/3-octave band. A step-function was included as a limit case for a wide-band signal meeting this requirement.

The mean data from the four subjects are presented in Fig. 6. for the white-noise masker and in Fig. 7. for the pink-noise masker. The dashed lines

represent the $5\log(n)$ and $10\log(n)$ integration rules, respectively. A least-squares linear regression fit with $y = \alpha + \beta \log(n)$ applied to the first four data points, for the white noise and the pink noise individually, yields β -values of 7.3 and 8.7, respectively. For the pink-noise masker the intersection with the wide-band condition (step function with $(E/N_0)_{1/3} = 0.5$ dB) is at a bandwidth of 10.1 1/3-octaves.

In global terms the data can be described by a $8\log(n)$ integration rule, n being the number of 1/3 octave bands covered by the signal. This integration rule appears to apply to signal bandwidths up to about three octaves.

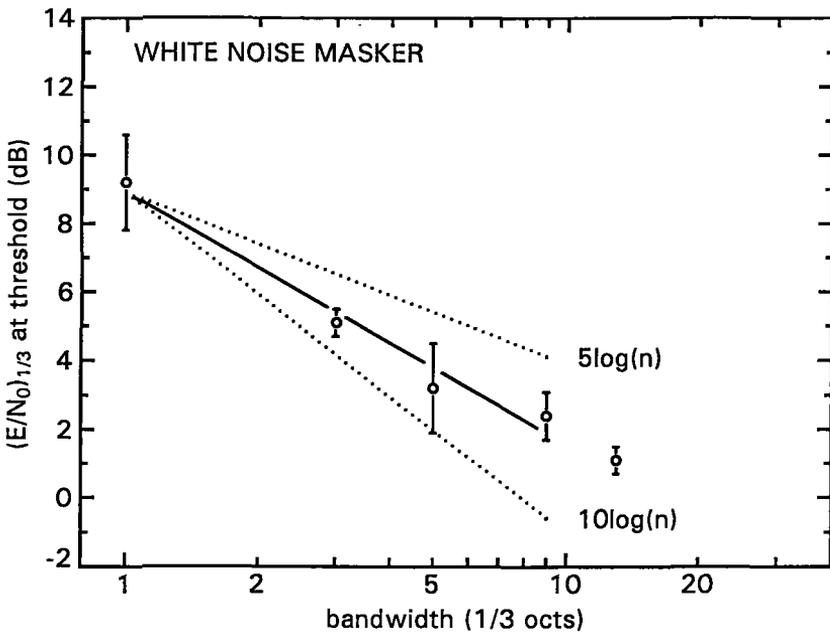


Fig.6. Masked threshold as a function of bandwidth of equal- $(E/N_0)_{1/3}$ complex signals (geometrically centered at 1.6 kHz), in case of a white-noise masker. The datapoints were extracted from psychometric functions and the mean and the standard deviations for the four subjects are indicated. The line represents the linear regression fit through four datapoints excluding the result for the largest bandwidth.

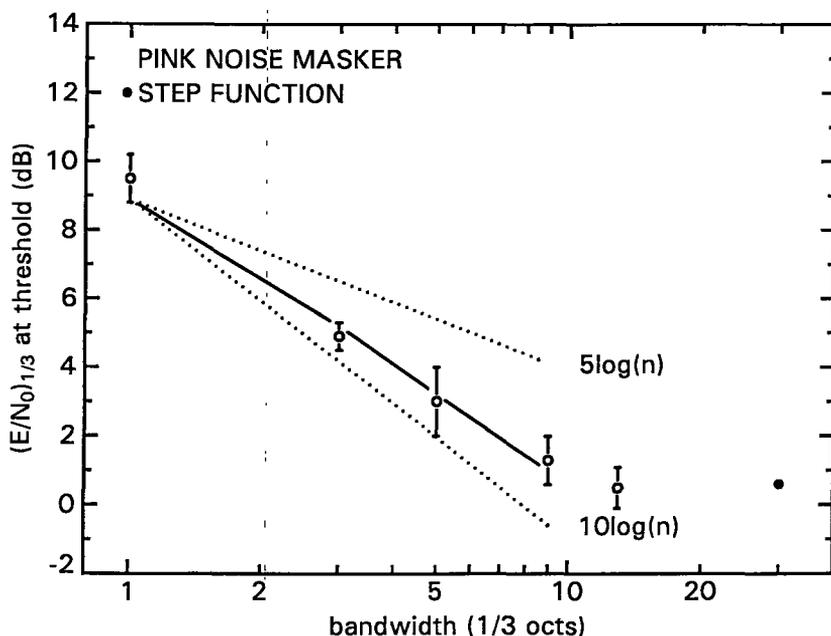


Fig.7. As Fig.6., for a pink noise masker. The detection threshold for the step function is included as a limit case towards large bandwidths.

1.4.2 Dirac pulses in white noise

To illustrate the significance of matching the signal spectrum to equal 1/3-octave detectability when investigating across-frequency integration, we considered Dirac pulses in white noise, being a very common stimulus in psychoacoustics. In this case, $(E/N_0)_{1/3}$ is not constant but increases with 1 dB for every next higher 1/3 octave band.

We considered Dirac pulses with increasing bandwidths, while fixing either the low or high cut-off frequency, rather than the center frequency as done in the previous experiment. Thus, one set of signals started with a low-frequency 1/3 octave band (centered at 630 Hz) while progressively more 1/3-octave bands were added at the high-frequency side, whereas the other set started with a high-frequency 1/3 octave band (4 kHz) while progressively more 1/3-octave bands were added at the low-frequency side.

The data are presented in Fig. 8. The masked threshold for the wide-band condition (nine 1/3-octave bands) serves as a baseline (0 dB), and the ordinate represents the required amplification of the Dirac pulse to reach

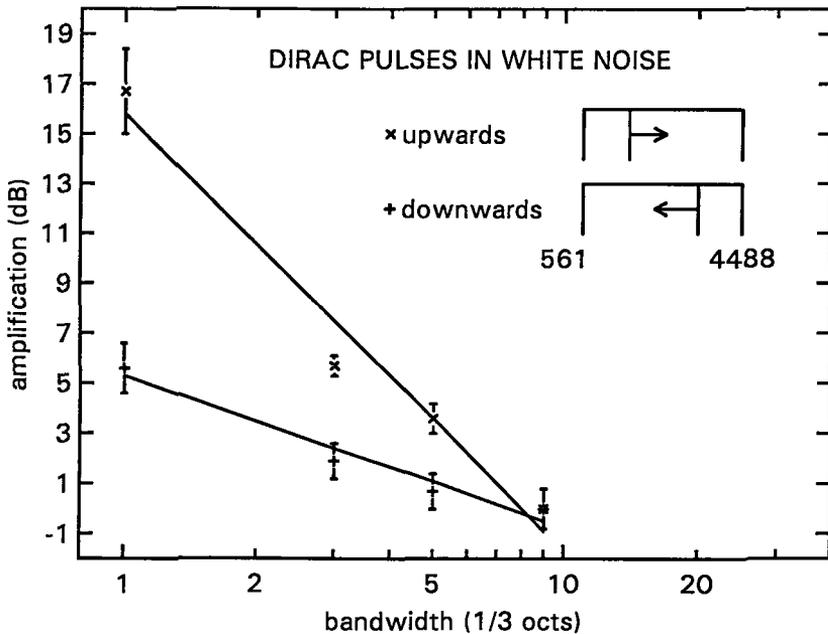


Fig.8. Required amplification to reach threshold relative to the wide band condition (9/3-octave bandwidth) as a function of bandwidth for filtered Dirac pulses with a fixed low cut-off frequency or a fixed high cut-off frequency, respectively. The masker was white noise. The lines represent the linear regression fits through the two sets of datapoints for which the mean and the standard deviation for the subjects are indicated. The datapoints were obtained with a three-down one-up adaptive level method yielding an estimate of the 79.4% correct score.

masked threshold for the various band-limited conditions. (In this case $(E/N_0)_{1/3}$ is not a relevant quantity for plotting the data, because it is not equal for the different 1/3-octave bands covered by the signal.)

A least square linear regression fit with $y = \alpha + \beta \log(n)$ applied to the datapoints yielded β -values of 6.0 and 17.5 for adding low-frequency bands and adding high-frequency bands, respectively. Indeed, as expected, the single low-frequency band (center frequency 630 Hz) requires more amplification than the single high-frequency band (center frequency 4 kHz). Consequently, these curves would lead to erroneous "integration rules" when interpreted simply as the effect of signal bandwidth on detection: for the upper curve the bandwidth effect would be strongly over-estimated (adding superior

high-frequency bands) and for the lower curve it would be strongly underestimated (adding inferior low-frequency bands).

1.4.3 Control experiment

A control experiment was included to verify that for longer duration broad-band signal indeed a $5\log(n)$ across-frequency integration rule is obtained when using the same experimental paradigm, for which the short duration signals yielded the $8\log(n)$ rule. A set of five compound gaussian signals, all with a geometrical center frequency of 1.6 kHz, was constructed with an α -value (Eq.1) of 0.00625. The effective duration $T_{\text{eff}} = (1/(\alpha f_0)) = 100$ ms for $f_0 = 1600$ Hz, with a -6 dB duration of 94.0 ms. The five compound gaussian signals had bandwidths of 1/3, 3/3, 5/3, 7/3 and 9/3 octaves. The three-down one-up adaptive level method was used which provides the 79.4% correct score, with separate adaptive tracks for the five signals being interleaved. A pink noise masker was applied and the signals were constructed to have equal $(E/N_0)_{1/3}$. For each of the four subjects ten threshold estimates for each signal were obtained and averaged.

The data are presented in fig.9. Linear regression with $y = \alpha + \beta \log(n)$ gives an β -value of 4.3 being close to the $5\log(n)$ across-frequency integration rule commonly found for long duration signals.

This control experiment shows that the efficient across-frequency integration for short duration signals as observed in our data is not merely an effect of the experimental paradigm used.

1.4.4 Non-contiguous signal spectra

This experiment was performed to investigate the spectral integration rule for non-contiguous signal spectra. For that purpose two sets of compound Gaussian signals were composed. The first, a set of four signals, consisted of the sum of three single Gaussian signals (each covering essentially a single 1/3-octave band, Sec. 1.2.1). Center frequencies were 1.25, 1.6, 2.0 kHz (contiguous 1/3-octave bands, see Fig. 2) or 1.0, 1.6, 2.5 kHz (spectral gaps of 1/3-octave) or 0.8, 1.6, 3.15 kHz (spectral gaps of 2/3-octave) or 0.63, 1.6, 4.0 kHz (spectral gaps of 3/3-octave). The second, a set of two signals, consisted of the sum of five single Gaussian signals, with center frequencies of 1.0, 1.25, 1.6, 2.0, 2.5 kHz (contiguous 1/3-octave bands) or 0.63, 1.0, 1.6, 2.5, 4.0 kHz (spectral gaps of 1/3-octave). Thus, all signals had a common (geometrical) center frequency of 1.6 kHz, and signal bands were added symmetrically at both the low- and high-frequency side. For both sets the

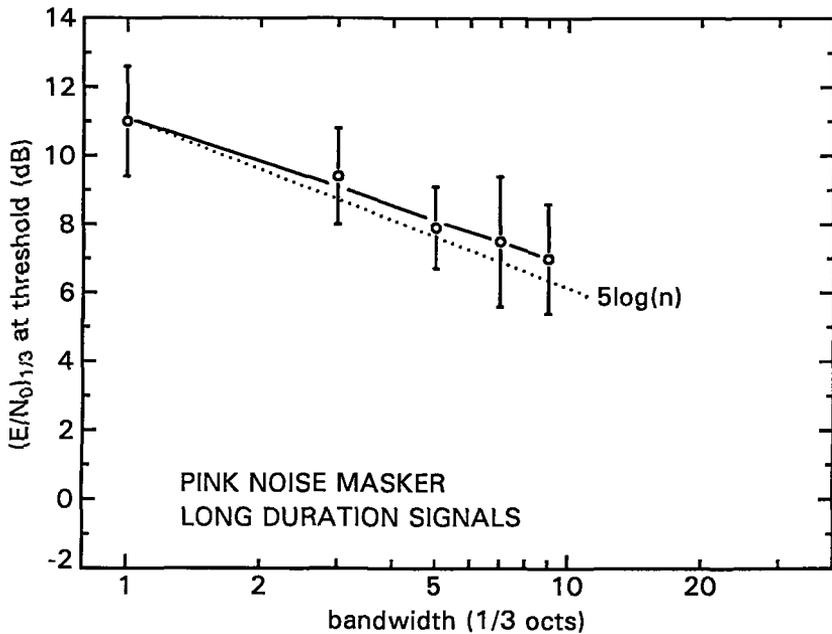


Fig. 9. Masked threshold as a function of bandwidth for long duration signals. The datapoints were obtained with a three-down one-up adaptive level method yielding an estimate of the 79.4% correct score. The line represents the linear regression fit through the datapoints for which the mean and the standard deviation are indicated. The dashed line represents the $5\log(n)$ integration rule.

maximum spectral range was three octaves wide, within the integration bandwidth for signals covering contiguous 1/3-octave bands. The masker was pink noise and $(E/N_0)_{1/3}$ was equal for all 1/3 octave bands excited by the signal.

The results are presented in Fig. 10. As a reference the data points in case of contiguous signal spectra (Fig. 7) are included. It should be noted that these signals were defined differently: namely as complex signals with, by definition, a smooth spectrum (Fig. 3 presented an example). The close agreement between the data points for common conditions (at spectral ranges of three or five 1/3-octaves) suggest that, indeed, $(E/N_0)_{1/3}$ is a relevant measure for detectability of broad-band signals, irrespective of signal definition. The data points referring to the first set (three 1/3-octave bands) and to the second set (five 1/3-octave bands) both show a slight increase of the

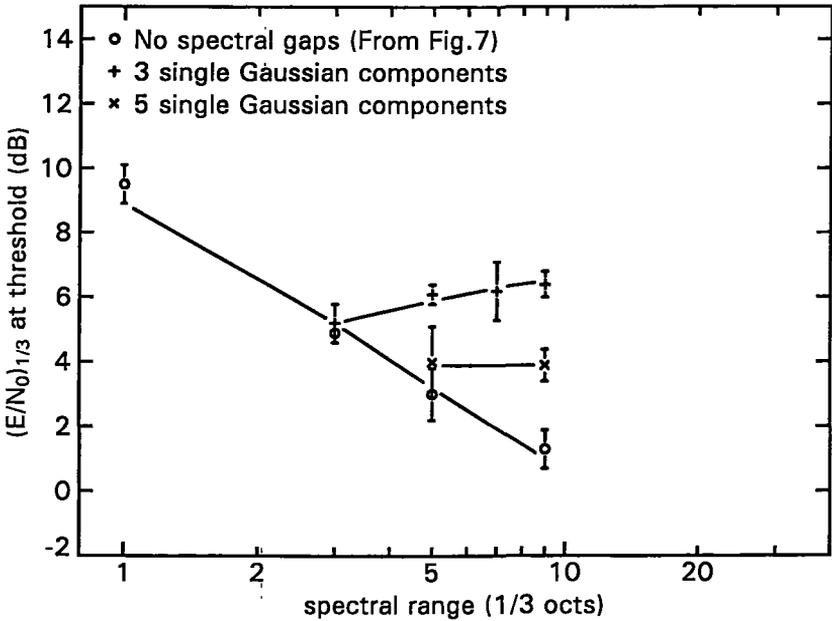


Fig.10. Detection threshold as a function of the spectral range covered by the contiguous and noncontiguous signal spectra. The open circles (contiguous spectra) represent data from Fig.7. The datapoints were extracted from psychometric functions and the mean and standard deviation are indicated.

masked threshold with increasing spectral range or, in other words, with increasing width of the spectral gaps between the single Gaussian signals. However, the effect is very small: for the first series, the maximum difference (i.e., three single Gaussian signals adjacent within one octave or distributed over three octaves) amounts to only 1 dB.

Thus, to a first-order approximation, the $8\log(n)$ integration rule found for signals with contiguous spectra also applies to signals with noncontiguous spectra (within a spectral range of three octaves).

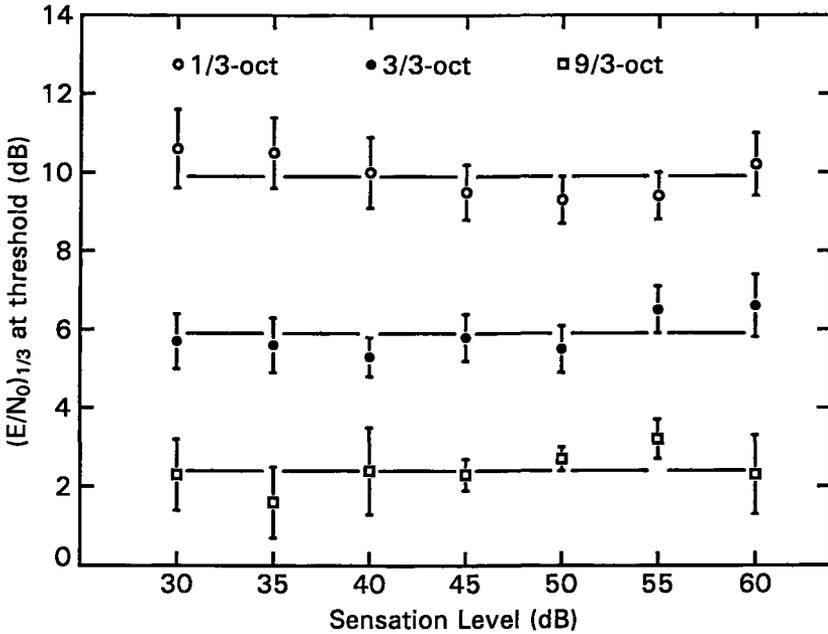


Fig.11. Detection threshold as a function of masker level for complex signals with bandwidths of 1/3, 3/3, and 9/3 octaves. The three lines represent the mean of each set of datapoints for which the mean and the standard deviation are indicated. The datapoints were obtained with a three-down one-up adaptive level method yielding an estimate of the 79.4% correct score.

1.4.5 Stimulus level

Some recent data on signal detection (Festen and Dreschler, 1988) show a peculiar interaction between bandwidth and stimulus level. This initiated the present experiment, investigating the robustness of the $\log(n)$ rule for spectral integration at different stimulus levels. Pink noise was used as the masker, at sensation levels ranging from approximately 30 dB SL up to 60 dB SL, in steps of 5 dB. These sensation levels correspond to a spectral density at 1.0 kHz ranging from approximately 5 up to 35 dB/Hz. Three complex signals were used, with bandwidths of 1/3, 3/3 and 9/3 octaves and equal- $(E/N_0)_{1/3}$ (the same signals for which data for the pink noise masker at 50 dB SL were presented in Fig. 7). The data are presented in Fig. 11. Over the present range of stimulus levels, no systematic effects are observed. An analysis of variance

showed no significant effect of sensation level on the masked threshold and no interaction between masker level and type of signal.

This indicates essentially no effect of sensation level: in Fig. 11 the three sets of data points have been fitted by horizontal lines, representing the mean across each of the three sets of datapoints, which are at distances in agreement with the $8\log(n)$ integration rule.

1.5 DISCUSSION

Below, two models will be considered which have been used to describe the masked threshold of multi-component signals. The first one, the statistical summation model (Green, 1958), states that the d_N' of an N -component signal (d_N') is composed out of the d -primes of the individual components (d_i') according to

$$d_N' = \sqrt{\sum_{i=0}^N d_i'^2} . \quad (2)$$

The second one, the independent probability model (Zwicker, 1967), states that the probability of not detecting an N -component signal ($1 - p_N$) is related to the probabilities of not detecting any of the individual components ($1 - p_i$) according to :

$$(1 - p_N) = \prod_{i=0}^N (1 - p_i) . \quad (3)$$

In case of equal detectability of the individual components we refer to Buus et al. (1986) the statistical summation model (eq.2) then reduces to :

$$d_N' = d_1' \sqrt{N} . \quad (4)$$

Thus, for each doubling of n the value of d_1' can be decreased by a factor $\sqrt{2}$ in order to yield a constant value for d_N' . Since $d_1' = 0.1 (E/N_0)$ the masked threshold for the compound signal is reached when for each doubling of n , the energy E of the individual components is reduced by 1.5 dB. This amounts to a $5\log(n)$ integration rule.

The independent probability model (eq.3), in case of equal detectability of the individual components, reduces to :

Thus, in principle, given the psychometric function for one individual

$$(1-p_N) = (1-p_1)^N . \quad (5)$$

component, p_1 as a function of signal level can be derived; from that, p_N as a function of signal level can be obtained from the above equation (eq.5) and the masked threshold of the compound signal can be defined as the signal level at which $p_n = 0.5$.

This procedure relies on the precise shape of the psychometric function for a single component, and especially, on the statistically unreliable region where the 2-IFC score approaches the 50% chance level. For instance, the threshold level of a 10-component signal should be derived from the point on the psychometric function where the score reaches 53.3%. The limited degree of accuracy of our psychometric functions in that region did not allow us to apply this procedure to our data. For a wideband signal (24 independent bands) Zwicker and Feldtkeller (1967) have shown experimentally that the masked threshold is reached when the signal level in the individual bands is about 6 dB below the masked threshold of each individual band, and that this fits the predictions of the independent probability model. Since $4.3 \log(24) = 6$, this would indicate an integration rule of typically $4\log(n)$.

We will briefly consider the fact that by our choice of equal $(E/N_0)_{1/3}$ signals the condition of equal detectability is actually not fulfilled exactly, but only approximated (Sec. 1.3, Fig. 5): for constant $(E/N_0)_{1/3}$ the d-prime of the individual 1/3-octave components are not the same but decrease slightly towards higher frequencies. As an example we will consider the case of broadening signal bandwidth around 1.6 kHz (Fig. 6. and Fig. 7.). For the statistical summation model, the effect of the d-primes being not equal can be calculated by applying the original eq.2 instead of eq.4. It is found that the effect is small. For instance, increasing the number of bands from 1 to 9 yields a threshold shift of 5.3 dB, whereas when assuming equal d-primes (eq.2), the predicted threshold shift is 4.8 dB. It may be concluded that the statistical summation model, also when considering the slight deviation from equal detectability, does not predict the effect of bandwidth on the detection threshold as observed in the present data.

The present data, indicating efficient across-frequency integration characterized by the $8\log(n)$ rule, are not in line with the models discussed above. It has been shown that these models (4 to $5 \log(n)$) describe across-frequency integration for multi-component signals with a duration of typically 100 ms or more, whereas the present data were collected for very brief signals. Thus, as Scholl (1962) has already shown for noise signals, signal

duration appears to be an important factor with respect to across-frequency integration, also in case of deterministic signals.

1.6 CONCLUSIONS

- Brief deterministic signals are effectively integrated across frequency bands: the masked threshold level, typically decreases with $8\log(n)$ dB, n being the number of 1/3 octave bands equally excited by the signal (Fig. 6 & Fig. 7).
- In order to investigate across-frequency integration care must be taken to match the signal level within each excited 1/3-octave band to ensure equal detectability across all relevant 1/3-octave band (Fig. 8).
- The across-frequency integration rule for short duration signals also applies to non-contiguous signal spectra covering a total spectral range up to three octaves (Fig. 10).
- The across-frequency integration rule for short duration signals remains intact over at least a 30 dB range of masker levels (Fig. 11).
- The present ($8\log(n)$) integration rule for deterministic short duration signals is not in line with current models on multiple component signal integration, while these models have been shown to hold for long duration signals (4 to 5 $\log(n)$ for durations of 100 ms or more).

Acknowledgement

This investigation was supported by the Netherlands Organization for the Advancement of Pure Research (N.W.O.). We want to thank Reinier Plomp, Guido Smoorenburg and Joos Vos for critically reading earlier drafts of the manuscript.

CHAPTER 2

SPECTRO-TEMPORAL INTEGRATION IN SIGNAL DETECTION

Willem A.C. van den Brink and Tammo Houtgast

Journal of the Acoustical Society of America 88, 1703-1712 (1990).

ABSTRACT

This paper is concerned with aspects of temporal integration and across-frequency integration in signal detection. Previous experiments on the detection of brief broadband signals (clicks) in continuous broadband noise revealed efficient spectral integration. The extent to which this effect is restricted to a critical time window was investigated by manipulating the temporal relations among the signal components in different frequency regions. In a typical experiment, the signal consists of nine brief Gaussian-shaped tone pulses, equally distributed at 1/3-oct intervals, each with a spectral width of about 1/3-oct, and each equally detectable in white noise. In the synchronized condition (i.e., coinciding peaks of the nine Gaussian envelopes), the detection threshold is reached when the levels of the nine individual tone pulses are about 8 dB below their individual threshold levels (efficient spectral integration). When the signal is progressively desynchronized (i.e., non-coinciding peaks of the Gaussian envelopes), detection threshold is found to increase. This suggests that efficient spectral integration in signal detection is confined to a narrow time window, with a typical value of 30 ms. Similar experiments were performed with respect to the efficiency of temporal integration. For constant-duration signals (100 ms) the detection threshold is found to increase when progressively widening signal bandwidth. The data indicate that the efficient temporal integration in signal detection is confined to a narrow frequency window which, not surprisingly, corresponds to the critical bandwidth.

2.1 INTRODUCTION

This paper is concerned with the effect of the spectro-temporal pattern of a signal on its threshold in continuous broadband noise. When considering integration of signal information in detection experiments, two special cases can be distinguished. One special case involves the temporal integration of narrowband signals. The classical experiments on temporal integration (e.g., Green et al., 1957; Plomp and Bouman, 1958; Zwislocki, 1960) revealed that the masked threshold of these signals is determined by their total energy, with an integration time of about 100 to 200 ms. The other special case involves spectral integration of short-duration signals. Scholl (1962) using bandpass-filtered noise bursts, and van den Brink and Houtgast (1990) using brief (typically 10 ms) deterministic signals, showed efficient spectral integration; for brief signals in a white-noise masker it is, essentially, the total signal energy which determines the masked threshold, irrespective of signal bandwidth (from 1/3 oct to 3 octs). However, for signals with spectro-temporal patterns deviating from these two special cases (i.e., narrowband temporal integration and short-term spectral integration), the integration of information is less effective.

Data obtained by Buus et al. (1986) and also by van den Brink and Houtgast (1990) indicate that spectral integration is less effective for long-duration signals (about 100 ms) than for brief (10 ms) signals. Typically, when widening the bandwidth of long-duration signals in white noise the total energy required for detection increases beyond a bandwidth of 1/3-oct, whereas the total energy required for detection of a 10 ms signal is essentially constant when widening bandwidth. Presently, only limited data are available with respect to temporal integration for broadband signals. Data collected by Scholl (1962), using bandpass-filtered noise bursts, and Nabelek (1978), using upward and downward tone glides, indicate that temporal integration for broadband signals is not as effective as for narrowband signals.

Related, but fundamentally different, are studies concerned with spectro-temporal resolution. Typically, the signal is relatively simple (click, tone) and the spectro-temporal pattern of the masker is manipulated (gaps in time or frequency, restricted bandwidths). For instance, the phenomenon of the critical masking interval was introduced by Penner et al. (1973) as a temporal analog to both Fletcher's (1940) and Greenwood's (1961a and 1961b) paradigm, with respect to the critical band. Recently Moore et al. (1987) performed a series of experiments with respect to the shape of the ear's temporal window. These experiments lead to a measure of auditory *resolution*, i.e., the ability of the auditory system to separate a masker and a signal in the spectro-temporal

domain, whereas the present study is concerned with integration of information across frequency and time.

This paper presents a systematic experimental study on the influence of the spectro-temporal signal pattern on the detection threshold in continuous broadband noise. The two main questions that will be addressed are concerned with, on the one hand, temporal integration as a function of signal bandwidth and, on the other hand, spectral integration as a function of signal duration. Pilot data, for a number of similar experimental conditions, were presented earlier by Houtgast (1987).

2.2 STIMULI AND PROCEDURE

2.2.1 Signal definition

Four types of signals were used in the present experiments, which will be referred to as single Gaussian, compound Gaussian, complex with group delay and bandpass impulse responses, respectively.

1. Single Gaussian

These signals consist of a tone pulse with a Gaussian envelope, defined by

$$s(t) = C \sin(2\pi f_0(t-t_0)) \exp(-\pi(\alpha f_0(t-t_0))^2) \quad (1)$$

where f_0 is the center frequency of the tone pulse. The value of a determines the effective duration ($t_{\text{eff}} = 1/(\alpha f_0)$) and the effective bandwidth ($b_{\text{eff}} = \alpha f_0$) of the signal. We used $\alpha = 0.2$ for short-duration signals for which the bandwidth is just within a 1/3-oct, and $\alpha = 0.00625$ for long-duration signals. For a single Gaussian signal with $f_0 = 500$ Hz and $\alpha = 0.2$ the -6 dB duration is 9.5 ms.

2. Compound Gaussian

The compound Gaussian signals are obtained by adding single Gaussian signals with different center frequencies, typically at 1/3-oct intervals, with the amount of desynchronization (relative temporal shift) of the individual Gaussian signals as a parameter.

3. Complex with group delay

These signals are defined in the time domain, with frequency components at 8-Hz intervals and a frequency-dependent phase shift in order to obtain group delays, linear on a $\log(f)$ scale. The signals are defined by the following equation :

$$s(t) = \sum_{n=70}^{561} \frac{\cos(2\pi n f_0 t + \varphi_n)}{\sqrt{n}}, f_0 = 8 \text{ Hz}. \quad (2)$$

The one-over-squareroot-n amplitude characteristic ensures that in each 1/3-oct band covered by the signal, the signal energy is constant. Thus, in a white noise masker, $(E/N_0)_{1/3}$ is constant. In case of a pink noise masker one-over-n amplitude characteristic is required. (See also Section 2.2.2).

With the phase $\varphi_n = k n f_0 (\ln(n) - 1)$, the group delay is given by $\tau_n = [d\varphi_n/d(nf_0)] = k \ln(n)$, where n represents the harmonic number of a stimulus component. Thus, the group delay is proportional with $\ln(n)$. For instance, for $k = (1000 \ln(2))^{-1} = 0.0014$, the delay amounts to 1 ms/oct. The signals were calculated for a 125 ms duration at 40- μ s sample intervals. One signal was calculated for a 250 ms duration at 40- μ s sample intervals. The effective signal duration is determined by the parameter k . Given the bandwidth of the signal (3 octs), the resulting signal envelope is approximately flat over a time interval of $1000 k \ln(8)$ ms. For the 125-ms duration signals, the value of k never exceeded 0.0462 (maximum group delay 32 ms/oct), so the flat signal envelope extended over maximally 96 ms. (k was 0.092 for the 250-ms signal, so the flat signal envelope extended over maximally 192 ms.) Beyond that interval, the envelope quickly drops to a much lower value (essentially random-phase addition). The onset and offset of the signals presentation were always in this low part of the signal envelope. These onsets and offsets were gated with a cosine-square window with a 5 ms rise and decay time.

4. Bandpass impulse responses

These signals are defined in the frequency domain at 5-Hz intervals. The amplitude characteristic was set to $1/\sqrt{f}$, this was done in order to obtain equal $(E/N_0)_{1/3}$ values for every 1/3-oct band stimulated by the signal, in case of a white noise masker (See also section B). The phase characteristic was set to zero, in order to obtain a linear phase impulse response. A 2048-point FFT was applied to obtain the time signal. The effective duration depends on the bandwidth. For a 1/3-oct bandwidth around 1.6 kHz the -6-dB duration is about 3.2 ms.

2.2.2 Level definition

The energy spectra of all signals were determined using a real-time frequency analyzer (Brüel & Kjaer 2123). The masking noise was defined by the spectral density N_0 , the mean noise energy within a time-frequency window of which the product of bandwidth and duration equals 1 or, equivalently, the mean noise intensity in a 1-Hz interval. Two types of masking noise were used: white noise (N_0 constant over f) and pink noise (N_0 decreasing in proportion to $1/f$). In all the experiments the level of the masking noise was set at approximately 50 dB SL, corresponding to a noise spectral density at 1.6 kHz of about 20 dB/Hz.

The levels of the signals were defined in terms of the energy in 1/3-oct bands, normalized by the local spectral density of the noise: $(E/N_0)_{1/3}$. For broadband signals, the $(E/N_0)_{1/3}$ -values in the various 1/3-oct bands covered by the signal were the same, thus giving approximately equal detectability of the individual signals within each of the 1/3-oct bands (van den Brink and Houtgast, 1990). The masked thresholds of the signals are expressed in terms of summated $(E/N_0)_{1/3}$, which is labelled $(E/N_0)_{tot}$. Thus, $(E/N_0)_{tot}$ represents the summation of the N_0 -normalized signal energy across the spectrum. In the special case of a white-noise masker the $(E/N_0)_{tot}$ equals E_{tot}/N_0 , with E_{tot} being the total, broadband, signal energy.

2.2.3 Apparatus

The experiments were performed under control of a PDP 11/34 computer. Signals were generated with a digital signal processor based on a TMS32010, at 40- μ s sample intervals with 10 kHz low-pass filtering. Broadband noise was generated, using a Wandel & Golterman RG-1 analog noise generator, and was digitally mixed with the signals. The stimuli were presented diotically (same stimulus to both ears) by means of Beyer DT-48 headphones. Subjects were seated in a soundproof room.

2.2.4 Subjects

Four subjects participated in the experiments. Two were experienced in psychoacoustic experiments, the other two were students who were familiar with auditory detection experiments, and were paid an hourly wage. The students were trained half a day before data collection started.

2.2.5 Procedure

A 2-IFC procedure was used. Each trial consisted of two 300-ms observation intervals, separated by 200 ms, and the signal was presented with equal probability in the first or second interval. Intervals were marked by a light on the listener's response box. Feedback was provided after each response. The broadband noise masker was presented continuously. In a series of trials signal presentation was randomized; i.e., for each trial, a signal was selected at random from the set of signals considered in one series (typically four to six different signals within one series). After each trial, the data for that signal were stored and a "new" signal was chosen at random. The signal level was varied adaptively. The separate adaptive tracks were interleaved in order to balance any possible effect of fatigue on the measured thresholds of the signals. Initially, the signal level was set well above masked threshold. The level was reduced by 4 dB after three successive correct responses. After the first incorrect response the stepsize was decreased to 2 dB. A three-down one-up procedure (Levitt, 1971) was applied in order to estimate the level that would produce 79.4% correct responses. After three reversals data collection started up to a total of thirteen reversals. Threshold was defined as the mean of the signal levels at the last ten reversals. One series, with four to six interleaved adaptive runs, took about 15 to 20 minutes. Ten such adaptive runs were obtained for each stimulus condition (in different series), and the threshold levels were averaged.

2.3 EXPERIMENTS

2.3.1 Two durations and two bandwidths

This baseline experiment was performed to verify the main point mentioned in the introduction, namely that broadband long-duration signals are less detectable (in terms of total energy) than narrowband and/or short-duration signals.

Four specific signal patterns (A, B, C, D) were constructed. The two narrowband signals were single Gaussian signals with a -values of 0.2 (B, narrowband short-duration) and 0.00625 (C, narrowband long-duration), with f_0 at 1.6 kHz. The two broadband signals were compound Gaussian signals consisting of nine single Gaussian components at 1/3-oct intervals with a geometrical center frequency of 1.6 kHz and with a -values of 0.2 (A, broadband short-duration) and 0.00625 (D, broadband long-duration),

respectively. No desynchronization was applied. Thus, the two narrowband signals had bandwidths of less than a 1/3 oct, while the two broadband signals covered 3-octs. The duration of the long-duration signals was a factor 32 longer than that of the short-duration signals. The masked thresholds of these signals were determined using a white-noise masker.

The results are presented in Fig. 1. The masked threshold of the broadband long-duration signal (D) is substantially higher than the masked thresholds for the three other conditions. This illustrates that temporal integration is less effective for broadband signals as compared to the narrowband signals, and also that spectral integration is less effective for long-duration signals as compared to short-duration signals.

For three of the signal patterns (narrowband and/or short-duration) the masked thresholds correspond to about the same amount of signal energy. The masked threshold data were subjected to an ANOVA [4(subjects) x 4(conditions) x 10(replica)]. The experimental conditions explained most of the variance in the data [$F(3,9) = 25.6$, $p < .0005$]. No significant interactions were found. A Newman-Keuls paired-comparison test revealed that condition D was significantly different from the other three conditions, whereas the other three conditions were not significantly different when compared to each other.

As mentioned above the data shown in Fig.1 were obtained using interleaved-adaptive tracks. It can be argued that by interleaving this method introduces an additional uncertainty in the detection task. The question that arises, is whether the poor integration as found for the broadband long-duration signals is an inherent property of the auditory system, or just a consequence of some sort of uncertainty or limitation in attention. It may be the case that, if such signals are mixed with brief and/or narrowband signals within a run, subjects may focus their attention on a narrow frequency region or a brief time interval.

To address this question, the masked thresholds of the four signals used in the previous experiment were also determined using a method in which the separate adaptive tracks were not interleaved. These measurements were performed with the same four subjects, and the series with only one signal were alternated with the interleaved-adaptive-tracks considered in the previous section.

The results obtained with both measuring methods (separate adaptive tracks interleaved or not-interleaved) was subjected to an ANOVA [2(methods) x 4(subjects) x 4(conditions) x 10(replica)]. The explained variance in the data due to the two different methods amount to 6.4 %, with $F(1,3) = 7.75$, $p >$

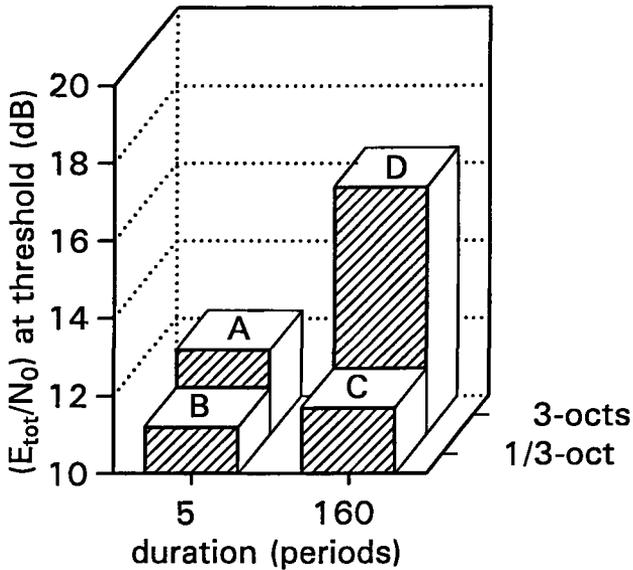


Fig.1. Masked threshold for four specific signal patterns, with two durations and two bandwidths. The masker was white noise.

0.07. The experimental conditions explained 61.1 % of the variance in the data, with $F(3,9) = 50.9$, $p < 0.0005$. No effect of replication was found in the analysis of variance, indicating that there was no learning effect. No significant interactions were found.

These results indicate that the differences between the masked thresholds of the four signals do not depend on the type of method used.

The data collected in the following experiments were also, for every experiment separately, subjected to an ANOVA, with [4(subjects) x (number of signal conditions) x 10 (replica)]. The results of all the analysis of variance showed that most of the variance was explained by the experimental conditions. No significant effect of replications was found (no learning effect). Also, no significant interactions were found (allowing simple averaging over subjects and replica).

The error bars in all figures indicate standard deviations, based on the four subjects and ten replica.

2.3.2 Restricted spectro-temporal surface

In the previous experiment it was found that the masked thresholds for three of the four signal patterns were reached at about the same amount of signal energy. An experiment was performed with these three signal patterns, with large differences in duration and bandwidth, included in a set of signals defined by one parameter: six bandpass impulse responses were constructed, all with a geometrical center frequency of 1.6 kHz, and with bandwidths of 3, 1, 1/3, 1/9, 1/27 and 1/81 octs. Since the effective duration is inversely proportional to the bandwidth, the spectro-temporal surface covered by each of these signals is the same, while the spectro-temporal shape varies systematically from broadband short-duration to narrowband long-duration. The masked thresholds of these signals were determined using a white-noise masker. The results are plotted in Fig. 2.

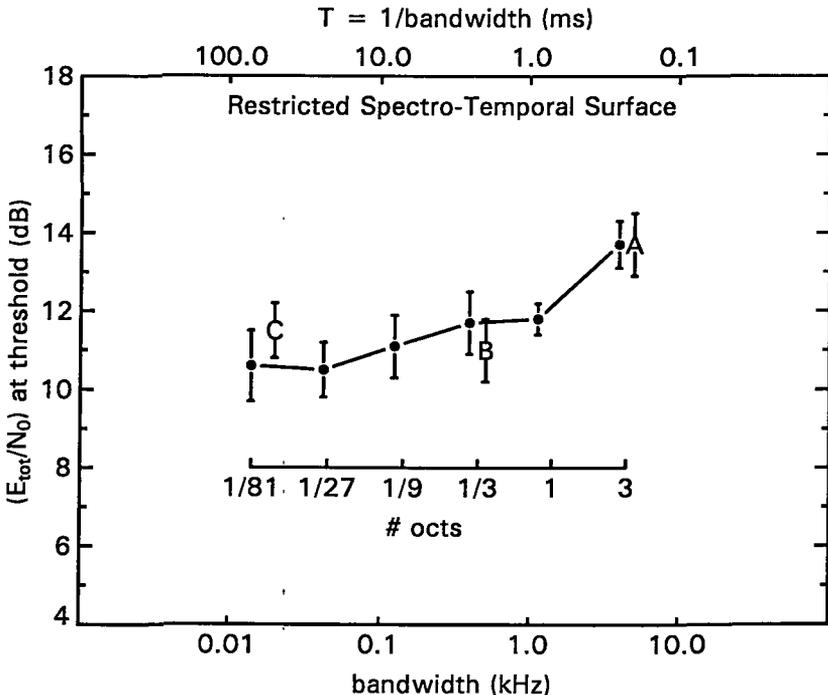


Fig.2. Masked threshold, as a function of bandwidth, for signals (bandpass impulse responses) all occupying the same spectro-temporal surface geometrically centered around 1.6 kHz. The masker was white noise. The results for the signal conditions A, B and C from Fig.1 comparable with bandwidth conditions of 3, 1/3 and 1/81 octs, respectively, have been included.

The bandwidths of 3, 1/3 and 1/81 octs are comparable with the three specific cases of the previous experiment, which were marked A, B and C, and are included in Fig. 2. Over a wide range of bandwidths (and corresponding durations), total signal energy determines the masked threshold, with the tendency that for bandwidths increasing towards 3 octs the integration is somewhat less effective, requiring somewhat higher signal energy. A similar experiment was performed by Festen and Dreschler (1988) using Gaussian-shaped tone pulses, varying systematically from short-duration broadband to long-duration narrowband. The results of Festen and Dreschler are similar to the present data (i.e., the masked threshold is determined by the total signal energy), but they are not comparable in detail because their signals were not adapted for equal detectability across frequency.

This graph does not give any indication of a "critical bandwidth" or a "critical time window" for efficient integration of information. This issue will be addressed in the next series of experiments, by studying the effect of either widening signal bandwidth for the long-duration condition C ("critical bandwidth" for temporal integration) or increasing signal duration for the broadband condition A ("critical time window" for spectral integration).

2.3.3 Critical bandwidth for efficient temporal integration

An experiment was performed to investigate whether there is a critical bandwidth for efficient temporal integration. The signals used consisted of nine temporally successive single Gaussian components ($a = 0.2$) at 8 ms intervals and with a geometrical center frequency of 1.6 kHz. The frequency range covered by these signals is defined by the ratio of the successive center frequencies of the Gaussian components within a signal, which was chosen as either 1.014, 1.029, 1.059, 1.122 or 1.258. These frequency ratios correspond to a total spectral range covered by the compound signal of 1/6, 1/3, 2/3, 4/3 or 8/3 octs, with all Gaussian components equally spaced on a logarithmic frequency scale. For the reference signal, a ratio of 1.0 was used, i.e. the nine Gaussian components all had a center frequency of 1.6 kHz. Two approaches for implementing the frequency spread were chosen. In the first approach, the f_0 values of the nine successive Gaussian components were randomly distributed over the nine possible values within a frequency range. A set of six quasi-random sets was constructed for each frequency range. In the second approach a sweep-like distribution of f_0 was used with a constant increment (or decrement) of $\log(f_0)$ for the successive components. In all conditions a white-noise masker was applied.

The results for the random f_0 -distribution, which showed no systematic differences across subjects and random sets, were averaged and are plotted in Fig. 3. The results for the sweep-like f_0 -distribution, averaged across subjects, are plotted in Fig. 4. The masked threshold for the reference condition is higher than the masked threshold for the comparable condition C in Fig. 2. This may be due to the fact that the signal has now been split up into nine components, whereas the masked threshold in Fig. 2 was obtained using a signal that was continuous in time. It can be seen from Fig. 3 that the masked threshold increases as the bandwidth is increased from 1/3 oct up to 4/3 oct, after which it remains about the same. The results for the sweep-like f_0 -distributions, plotted in Fig. 4, indicate that when the spectral range covered by the signal is increased, the masked threshold remains about the same up to a spectral range of about 1/3-oct after which the masked threshold increases.

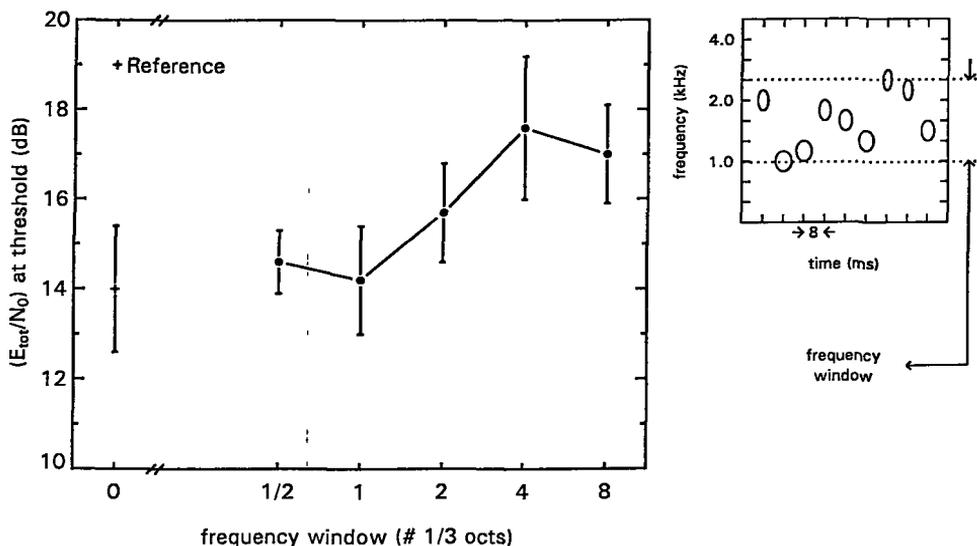


Fig. 3 Masked threshold of nine successive single Gaussian signal components as a function of the total frequency window covered by the signal. Within the frequency window covered by the signal the components were randomly distributed. The masker was white noise. An example of a spectro-temporal signal pattern within a frequency window is given on the right side of the data plot.

It should be noted that the spectral range covered by each of the nine signal components is about 1/3-oct, which actually restricts the spectral range conditions to a lower limit of a 1/3-oct. This is a consequence of the requirement that the nine Gaussian tone pulses have a total duration of less than 100 ms, which limits the tone pulse duration and subsequently the spectral range covered by each single tone pulse. So the present results (effect on masked threshold when widening bandwidth beyond 1/3-oct) may partly reflect the characteristic of our test signals. However, a comparison of our data with those of Nabelek (1978) suggests that the present data may be generalized. This is illustrated in Fig. 5, which shows data adapted from Nabelek (1978), who used upward and downward tone glides in a white-noise masker. The relative masked thresholds for tone glides with durations less than 100 ms have been plotted. To be able to make a comparison of our masked thresholds with Nabelek's masked thresholds, which are expressed in dB SPL, we have replotted all masked thresholds in terms of amplification relative to a reference signal. The reference signal in case of Nabelek's data consisted of the pure-tone masked threshold in the frequency region of the glide.

The dashed lines in Fig. 5 represent the data from Fig. 4. For downward tone glides the masked thresholds as found by Nabelek seem to be somewhat larger than those found in the current experiments. Although the data presented in Figs. 3 and 4. do not show a real breakpoint, they indicate that efficient temporal integration is limited to a certain bandwidth, with a magnitude of about 1/3 octave. Thus, in global terms the data indicate that temporal integration starts to deteriorate when signal bandwidth is increased beyond approximately 1/3 oct.

2.3.4 Critical time window for spectral integration

In the last series of experiments, the temporal analog to the previous experiment was performed. When desynchronizing the signal components (i.e., introducing a relative time shift), there are at least two choices as to how this can be performed.

On the one hand, the hypothetical temporal window might well be related to a certain fixed time interval, independent of frequency, suggesting that the various components should be shifted by a constant amount with respect to each other (a linear time grid).

On the other hand, it may be argued from a physical point of view that temporal resolution is inversely proportional to bandwidth. For the hearing

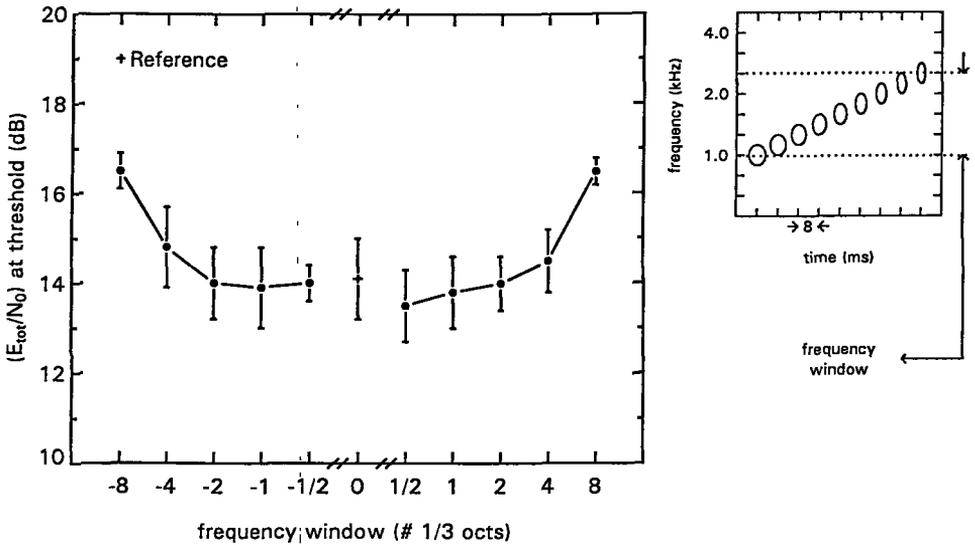


Fig.4. As Fig. 3, for a sweep-like distribution of the signal component frequencies. For positive values along the abscissa the high-frequency components are delayed (upward sweep), while for negative values (downward sweep) the low-frequency components are delayed.

system, this would amount to a $1/f$ proportionality, and the peaks of the envelopes of the signal components should thus be shifted in proportion to $1/f$, i.e. a larger shift should be applied for the low-frequency components than for the high-frequency components (a $1/f$ -proportional time grid).

A pilot experiment was performed in an attempt to discriminate between these two types of temporal desynchronization. Both types were applied separately to a compound signal consisting either of four low-frequency components (0.63, 0.8, 1.0, 1.25 kHz) or four high-frequency components (2.0, 2.5, 3.15, 4.0 kHz). Beside the reference condition, six conditions of desynchronization were constructed with increasing total time window. (The total time window is defined in terms of ms in case of the linear time grid, and in terms of periods in case of the $1/f$ -proportional grid.) For each condition, the four components within a signal were distributed at random. A set of ten quasi-random signals was constructed for each condition, for both types of desynchronization, and the four low and the high frequency

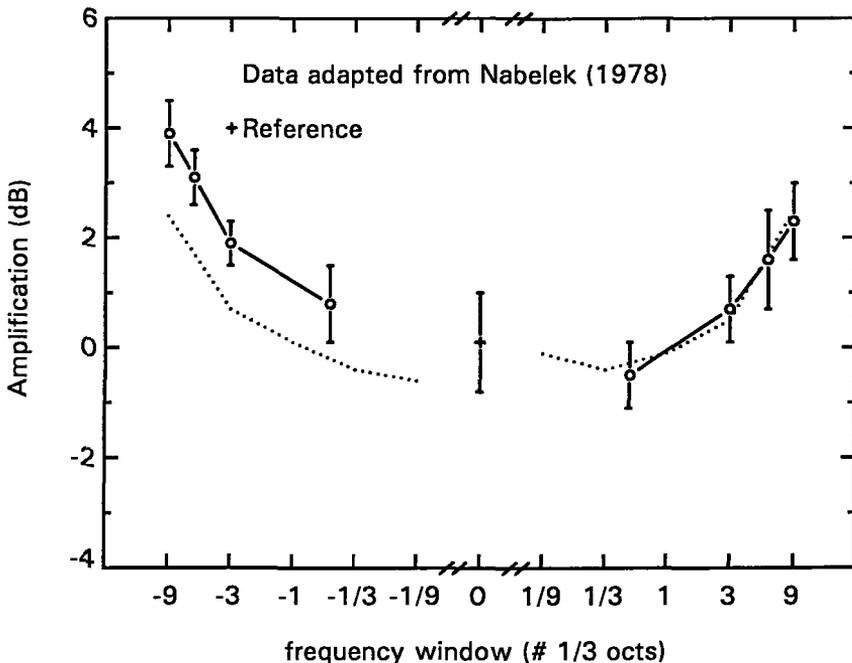


Fig.5. Data adapted from Nabelek (1978). Amplification to reach masked threshold as a function of the spectral range of the tone glide relative to the masked threshold of a pure tone in the glide region. The masker was white noise. The dashed curve represents the data from Fig. 4. For positive values along the abscissa the high-frequency components are delayed (upward sweep), while for neagtive values (downward sweep) the low-frequency components are delayed).

components. A pink-noise masker was applied in all conditions. That type of desynchronization which shows the same effect on the masked threshold for the low- and high frequency compound signals would appear to be the proper one to apply to the nine-component compound signal. Because no systematic differences were observed across the subjects and across the random sets the data were averaged. The data are presented in Figs. 6 and 7.

Although the differences are only small, the data in Fig. 6 indicate that, by and large, a time shift in ms has a larger effect for the high-frequency components than for the low-frequency components, whereas in Fig. 7, with a time shift proportional to $1/f$, the effects are not systematically different. This suggests that the $1/f$ -proportional time grid is to be preferred when investigating desynchronization effects over a wide frequency range.

Using this time grid, the temporal window for efficient spectral

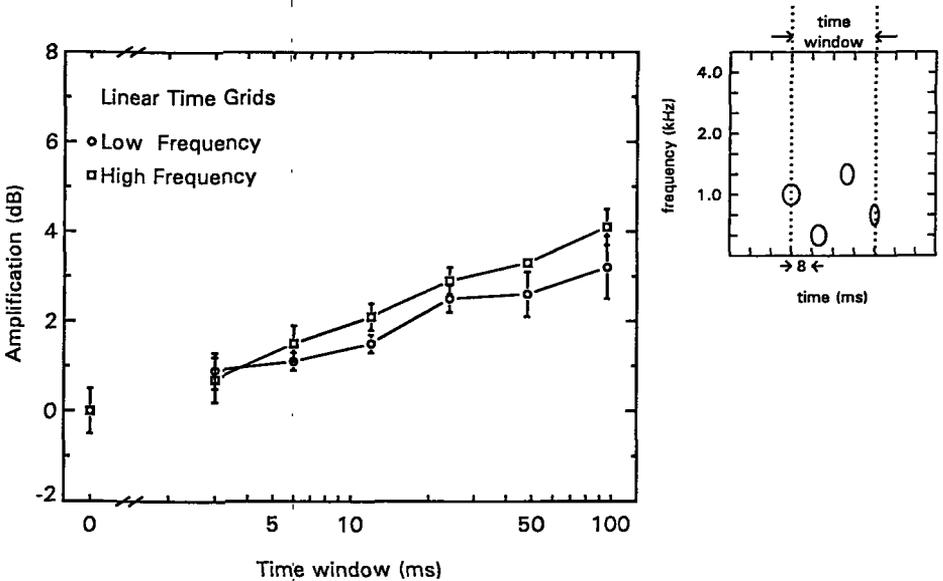


Fig. 6. Amplification, relative to reference (synchronized) condition, to reach masked threshold as a function of desynchronization in ms for two types of signals, each consisting of four single Gaussian components: a low-frequency set (at 0.63, 0.8, 1.0 and 1.25 kHz) and a high-frequency set (at 2.0, 2.5, 3.15 and 4.0 kHz). Within the time window covered the four signal components were chosen in a random order. The masker was pink noise. An example of a spectro-temporal signal pattern is given on the right side of the data plot.

integration was investigated with either a random or sweep-like distribution of the nine tone pulses within a window. A signal, typically, consisted of nine single Gaussian components ($a = 0.2$) with the f_0 -values at 1/3-oct intervals, geometrically centered around 1.6 kHz. In the reference condition, all the peaks of the envelopes of the signal components coincided in time. For the random distribution, a set of ten quasi-random distributions was constructed for each window. The results, averaged across random sets and subjects, are plotted in Fig. 8, in which the masked thresholds for the 1/f-proportional desynchronized four-component signals from the previous experiment have been included (mean data from Fig. 7). (In order to facilitate a comparison of the data for these different signal types, the data have been plotted in terms of amplification relative to the not-desynchronized (reference) condition.)

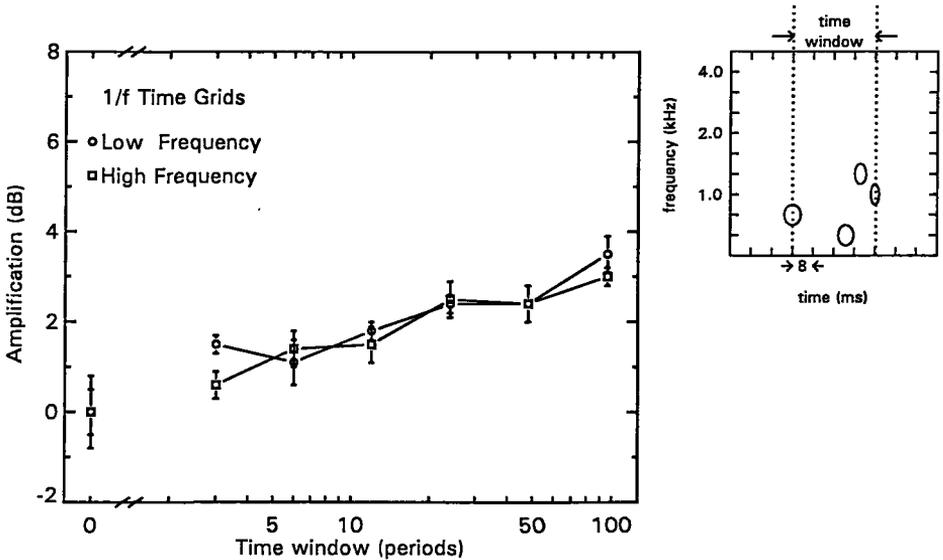


Fig. 7. As Fig. 6, using a desynchronisation in terms of $1/f$ (periods) rather than in ms.

For the sweep-like distribution of the nine f_0 values of the components within a signal, either with an increment or decrement between all the f_0 values within a signal, the results are plotted in fig.9.

For the latter type of desynchronization an asymmetric function is obtained for the masked threshold as a function of the time window (in periods). For small negative desynchronizations, for which the high-frequency components are leading within a signal, the masked threshold deteriorates more rapidly as compared to small positive desynchronizations, for which the low-frequency components are leading within the signal. This might be related to the travelling time of the signal along the basilar membrane: high-frequency fibers have smaller latencies than low-frequency fibers (e.g. van Heusden and Smoorenburg, 1981).

The data with the random distribution (Fig. 8) show a gradual increase of masked threshold with increasing time window, which would be expected when combining the effects of both positive and negative desynchronizations as observed in Fig. 9. These data can only be summarized in terms of a critical time window in a very global sense: that time window for which the

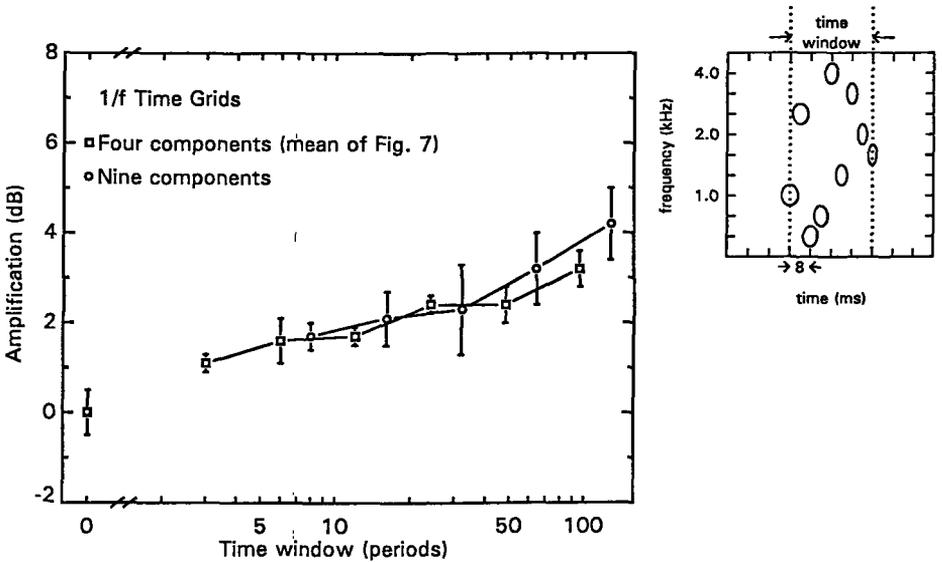


Fig. 8 Amplification to reach masked threshold, relative to the reference condition (synchronized), for 1/f-proportional desynchronization of a nine-component compound Gaussian signal, as a function of the total time window covered by the signal, with a random distribution of the signal components. The mean data from Fig.7 have been included. The masker was pink noise. An example of a spectro-temporal signal pattern within a time window is given on the right side of the data plot.

masked threshold is about half way between that for the condition of perfect synchrony and that for the condition of complete desynchronization, which amounts to about 40 periods. This would imply that the "critical" time shift between adjacent components within a signal (at 1/3-oct intervals) amounts to about 4 to 5 periods. However, this is about equal to the effective number of periods within each single Gaussian signal. Thus, it might be argued that the present result to some extent reflects the characteristics of the type of signal used. Therefore, an additional experiment was performed. Complex signals with group delays were constructed with a linear desynchronization on a log(f) scale introduced by a frequency dependent group delay (Sec.I A). For these complex signals with group delays the masked thresholds were determined as a function of the total time window covered.

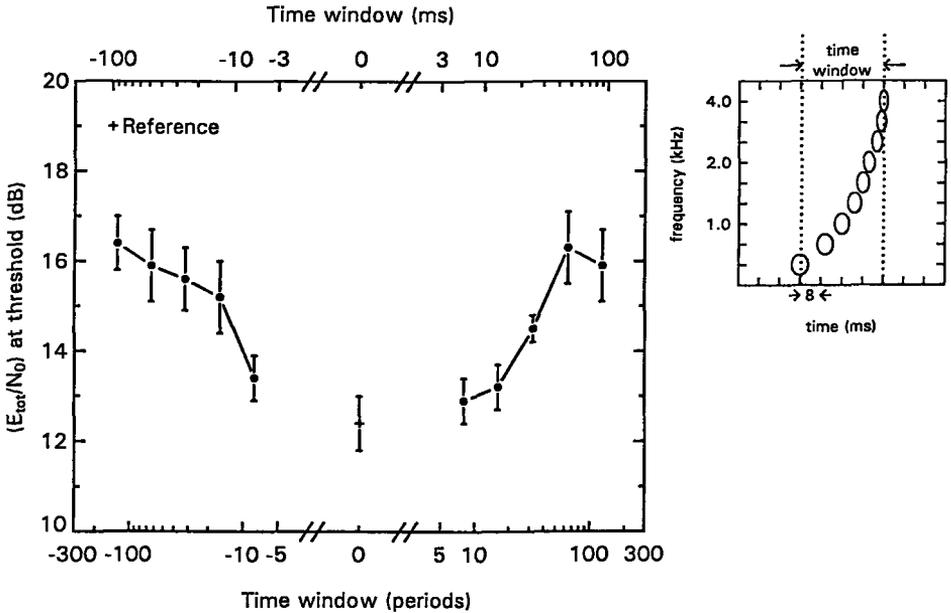


Fig. 9. Masked threshold for 1/f-proportional sweep-like desynchronisation of a nine-component compound Gaussian signal, as a function of the total time window covered by the signal. The masker was pink noise. An example of a spectro-temporal signal pattern is given on the right side of the data plot. For positive values along the abscissa the high-frequency components are delayed (upward sweep), while for negative values (downward sweep) the low-frequency components are delayed.

The results are plotted in Fig. 10. The data in Fig. 9 and Fig. 10 are very similar, indicating that the choice of signal is not critical when considering the deterioration of the masked threshold with increasing temporal window. In general terms, the "critical" width of the temporal window for spectral integration for these types of signals can be estimated at about 40 periods. For the present frequency range considered (3 octs with a geometrically center frequency at 1.6 kHz) this amounts to a duration of about 30 ms. But, for instance, for a 3-oct range a factor two higher in frequency, this would result in a temporal window of about 15 ms.

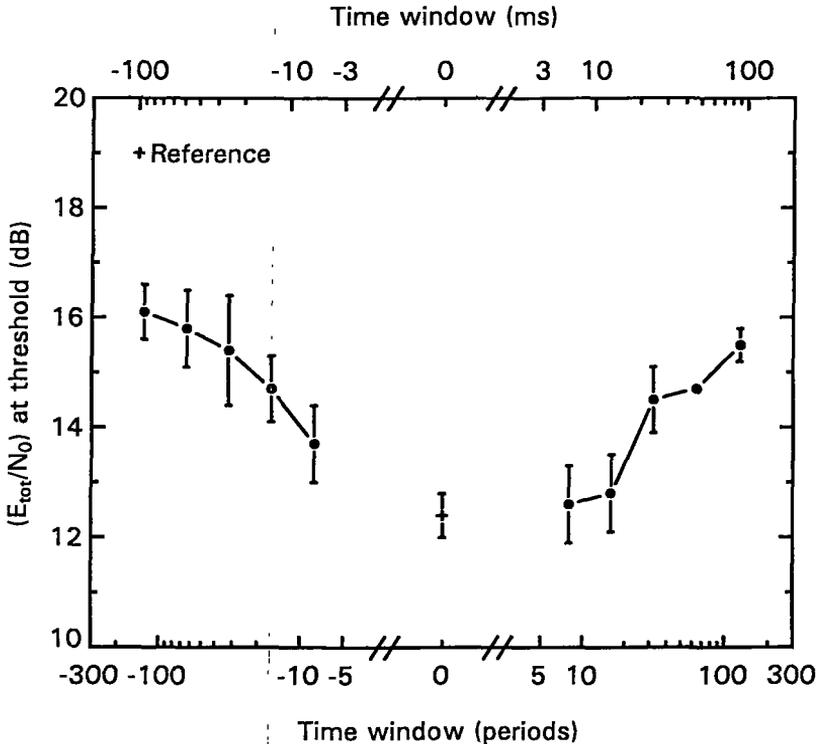


Fig. 10. Masked threshold for a complex signal (from 561 to 4488 Hz) with group delay, as a function of the total time window covered by the signal. The masker was pink noise. For positive values along the abscissa the high-frequency components are delayed (upward sweep), while for negative values (downward sweep) the low-frequency components are delayed.

2.4 GENERAL DISCUSSION

Throughout the paper we have distinguished between short-duration (about 5 periods) and long-duration (about 160 periods), and between narrowband ($\leq 1/3$ -oct) and broadband (3-octs) signals. This yields the four specific type of signals A, B, C and D considered in Fig. 1.

For B \rightarrow C (both narrowband, from short-duration to long-duration) we have verified the classical temporal integration rule: $(E/N_0)_{tot}$ at masked threshold is constant or, in other words, the SPL at masked threshold reduces by $10 \log(T_l/T_s)$ dB, T_l and T_s being the durations of the long-duration signal and short-duration signals, respectively. The transition from B \rightarrow A (both short-duration, from narrowband to broadband) reflects the almost equally effective spectral integration rule, as obtained in our previous study (van den

Brink and Houtgast, 1990): in terms of $(E/N_0)_{tot}$ type A requires 2 dB more energy at masked threshold than type B. Thus, within each 1/3-oct, the signal energy (normalized with respect to the "local" N_0) reduces as $8 \log(n)$ dB, n being the number of 1/3-octs covered by the signal.

The transition from C \rightarrow D (both long-duration, from narrowband to broadband) reflects the less efficient spectro-temporal integration for long-duration broadband signals: type D requires 6 dB more energy $((E/N_0)_{tot})$ at masked threshold than type C, corresponding with a spectral integration rule of $4 \log(n)$ rather than the $8 \log(n)$ rule for short-duration signals. The transition C \rightarrow D has been investigated in more detail: when gradually increasing the spectral range covered by the signal it is found that the spectro-temporal integration deteriorates for signal bandwidths beyond about 1/3-oct. Thus, the critical frequency window for efficient temporal integration amounts to about 1/3-oct. The transition A \rightarrow D (both broadband, from short-duration to long-duration) also reflects the less effective spectro-temporal integration for long-duration broadband signals: type D requires 4 dB more energy $((E/N_0)_{tot})$ at masked threshold than type A, corresponding with a temporal integration rule of $6 \log(T_1/T_2)$, rather than the classical $10 \log(T_1/T_2)$ rule for narrowband signals. The effect of the time window on efficient spectral integration was investigated by the same experimental paradigm as used for the "critical frequency window". When gradually increasing the width of the temporal window covered by the signal it is found that $(E/N_0)_{tot}$ at threshold is increased significantly when the time window exceeds 30 ms.

However, this value can not be generalized independently from the spectral content of the signal, since the time shift should probably be specified in periods rather than msec. Typically, the 'critical' delay between adjacent 1/3-oct bands is on the order of five periods. When integrating these delays over the spectral range covered by the signal (from f_1 up to f_2), this would lead to the following expression for the total time window in ms.

$$T_c = \left[\frac{5}{\ln(2)/3} \right] \left[\frac{1}{f_1} - \frac{1}{f_2} \right] [10^3] \text{ (ms)} . \quad (3)$$

For the present signal ($f_1 = 630$ Hz and $f_2 = 4000$ Hz), this corresponds to the critical time window of about 30 ms, as mentioned before. However, the equation above should not be taken too strictly; considering the underlying data (Figs. 8 to 10), the result given by Eq. (3) only provides a global indication of the critical time window involved in spectral integration.

The asymmetry of the effects observed for upward and downward sweep-like desynchronizations indicates that spectral integration is more resistant when the low frequency components lead within the signal. This may

be related to the physiological dispersion in which the low frequency signal components reach their place of stimulation a little later than the high frequency components.

2.5 SUMMARY AND CONCLUSIONS

Signal detection in continuous broadband noise was investigated as a function of the spectro-temporal pattern of the signal. These patterns were always restricted to a spectro-temporal region with a bandwidth of 3 octs (centered around 1.6 kHz) and a duration of 100 ms. The masker was white noise (N_0 is constant) or pink noise (N_0 proportional to $1/f$). Signals were designed with constant E/N_0 within each 1/3-oct band covered by the signal (constant $(E/N_0)_{1/3}$). Summation of these $(E/N_0)_{1/3}$ values is defined as the total N_0 -normalized signal energy: $(E/N_0)_{tot}$.

Essentially, two types of integration rule were found: efficient integration, for which the total N_0 -normalized signal energy determines the detection threshold, and a less effective integration for which each doubling of bandwidth or duration requires 1.5 dB more total N_0 -normalized signal energy at detection threshold. The efficient integration was observed for two spectro-temporal signal patterns: narrowband and/or short-duration. For a broadband and long-duration pattern, signal information is integrated less effectively. The efficient temporal integration (constant N_0 -normalized energy up to 100 ms) requires that the total bandwidth should not exceed 1/3-oct. The efficient spectral integration (approaching constant N_0 -normalized energy across 3 octs centered around 1.6 kHz) requires that the total temporal window covered by the signal should not exceed about 30 ms. This latter value depends on the frequency spectrum of the signal. When introducing sweep-like group delays across the signal spectrum an asymmetry was observed for downward or upward sweeps, which might be related to the physiological dispersion occurring in the inner ear.

Acknowledgements

This investigation was supported by the Netherlands Organization for Scientific Research (N.W.O.). We would like to thank Adelbert Bronkhorst for critically reading an earlier draft of this manuscript, and we would like to thank Brian C.J. Moore and an anonymous reviewer for their helpful comments.

CHAPTER 3

SIGNAL DETECTION IN TEMPORALLY MODULATED AND SPECTRALLY SHAPED MASKERS

Willem A.C. van den Brink, Tammo Houtgast and Guido F. Smoorenburg
Journal of the Acoustical Society of America (submitted)

ABSTRACT

The first part of this paper presents several experiments on signal detection in temporally modulated noise, yielding a more general approach towards the concept of comodulation masking release (CMR). We measured masked thresholds of both long- and short-duration narrow-band signals presented in a 100% sinusoidally amplitude-modulated (SAM) noise masker (modulation frequency 32 Hz), as a function of masker bandwidth from 1/3-oct up to 13/3 oct, while the masker band was centered at signal frequency. With the short-duration signals a substantial CMR was found, whereas for the long-duration signals CMR effect was smaller. Furthermore, we investigated whether CMR changes when the bandwidth of signals, consisting of bandpass filter impulse responses, is increased. The data indicate that substantial CMR remains even when all masker bands contain a signal component. The finding that substantial CMR remained while across-channel differences decrease is not in line with current models accounting for the CMR phenomenon. The second part of this paper concerns signal detection in spectrally shaped noise. We investigated whether release from masking occurs for the detection of a pure-tone signal at a valley or a peak of a simultaneously presented masking noise with a sinusoidally rippled power spectrum, when this masker was preceded and followed by a second noise (temporal flanker burst) with an identical spectral shape as the on-signal noise. Similar to CMR effects for temporal modulations, the data indicate that coshaping masking release (CSMR) occurs when the signal is placed in a valley of the spectral envelope of the masker, whereas no release from masking is found when the signal is placed at a peak of the spectral envelope of the masker.

3.1 INTRODUCTION

In this paper we attempt to integrate different across-frequency and across-time processing phenomena in auditory signal detection by manipulating the spectral and temporal parameters of both signal and masker.

A clear demonstration of the significance of across-frequency processing in auditory signal detection was shown by Hall and colleagues (Hall et al., 1984; Hall 1986). In a series of detection experiments they showed that masking of a pure-tone signal presented in a narrow-band modulated noise masker (masker band) around the signal frequency decreased, when a second narrow-band modulated noise masker (flanker band), remote in frequency from the first masker band, was added to the first one. This release from masking occurred only when the two noise bands had similar temporal envelopes. The phenomenon was called comodulation masking release (CMR). Most of the experiments addressing CMR were restricted to narrow-band signals and (multiple) narrow-band maskers.

Signal detection experiments in which unmodulated maskers (e.g. Green 1958, Buus et al. 1986, van den Brink and Houtgast, 1990) were used, revealed another type of across-frequency processing; across-frequency integration of signal energy. Across-frequency signal integration was shown to be more effective with short-duration signals than with long-duration signals.

CMR and across-frequency integration could simultaneously occur for broadband signals and maskers. The question of how these phenomena combine is the subject of the present study.

Some data about the effect on CMR of increasing the bandwidth of the signal are already available in the literature. Recently Hall and Grose (1988) reported a series of experiments on CMR, using maskers consisting of several narrow bands of noise (having similar or dissimilar temporal envelopes) and long-duration multi-component signals. They found that CMR is often smaller for multi-component signals than for single-component signals. When considering the experiments on the detection of multi-component signals in unmodulated noise by van den Brink and Houtgast (1990), it can be argued that the release from masking for multi-component signals, as presented by Hall and Grose (1988), is determined by not only CMR but also by across-frequency integration of signal energy.

It is also interesting to know what to expect from an increase in signal bandwidth according to the current models accounting for CMR. A number of models have been proposed in the literature to account for the CMR phenomenon. One explanation for CMR is that the auditory system performs an across-frequency comparison, for instance of the modulation patterns, of

the different frequency-channel outputs (Hall et al, 1984; Buus, 1985; Schooneveldt and Moore, 1987, 1989). When no signal is presented, the modulation patterns of the different frequency channels are similar. However, when a signal is presented in one of the frequency channels the amplitude modulation pattern of this channel differs from the amplitude modulation pattern of the other frequency channels, thus providing a possible detection cue. A second explanation was suggested by Buus (1985); he assumed that the amplitude modulation pattern in the frequency channels to which no signal components are presented (flanker bands) indicates the optimum moments in time to listen for the signal (the minima). Therefore this explanation has been named 'listening in the valleys'. A third explanation has been suggested by Cohen and Schubert (1985,1987) and by Richards (1987); CMR is thought to be closely related to the detection of a difference in correlation between the envelopes of the different frequency-channel outputs. Detection would occur when presentation of a signal causes the envelope correlation to fall below a criterion level. All present models described above suggest that no CMR is to be expected when a broadband signal is presented such that all frequency channels, containing comodulated masker energy, receive signal components with identical temporal envelopes.

One aim of the experiments, to be presented in the first part of this paper, is to separate across-frequency integration from CMR for detection of broad-band signals in a broad-band temporally modulated masker, both for short- and long- duration signals. In recent CMR experiments with short-duration signal (Grose and Hall, 1989; Moore et al.,1990) it was shown that CMR depends on the temporal position of the signal relative to the modulation cycle of the masker envelope. CMR was found only when the narrow-band short-duration signal was placed in the valley of the masker envelope; not at the peak. For a peak placement of the signal, masked thresholds increased when adding comodulated noise bands, which is referred to as across-channel masking (ACM). In view of these results we determined the masked thresholds for short-duration signals of various bandwidths placed at either the peak or the valley of the masker envelope.

The second part of the present paper focusses on the spectral analog of our CMR experiment. It addresses the question of whether release from masking occurs in the detection of a tone pulse simultaneously masked by a spectrally shaped noise burst, when an additional noise burst with an identical spectral shape as the on-signal noise precedes and follows the first masker (temporal flanker burst).

Penner and Cudahy (1973), using unshaped broadband noise, showed

that adding noise beyond the "critical masking interval" (10-15 ms) does not change the masked threshold of a brief broadband click. Because our masker durations always exceeded this "critical masking interval", no change in the masked threshold of our signals in the unshaped noise condition, when increasing masker duration, is expected. Other experiments (Viemeister, 1980; Carlyon, 1987, 1989) do show a release of masking for the detection of a tone pulse when adding a second masker, remote in time, from the first masker and the signal. In recent experiments Carlyon (1989) showed that the masked threshold of a brief tone pulse simultaneously masked by a noise burst containing a spectral notch around the signal frequency, was reduced by prior exposure to a 200 ms noise burst that had the same spectral shape as the on-signal masker. Carlyon's results were interpreted as a form of processing which groups together energy in frequency regions containing common temporal amplitude envelopes, and which enhances the internal presentation of newly-arriving energy in previously unstimulated frequency regions.

In summary, the second part of this paper presents a series of experiments with a stimulus paradigm similar to the one used in the CMR study, but with the time and frequency axis being interchanged. Thus, we investigated whether, similar to a CMR for signal detection in temporally modulated noise, a coshaping release from masking (CSMR) can be found for signal detection in spectrally shaped noise when nonsimultaneous coshaped maskers are added to the stimulus.

The different stimulus configurations are summarized schematically in Fig.1, in order to illustrate the similarity between these paradigms.

3.2 METHODS

3.2.1 Procedure

A two-interval forced choice procedure was used in all experiments. Each trial consisted of two 300-ms observation intervals, separated by 200 ms. The signal was presented with equal probability in the first or second interval. Intervals were marked by a light on the listener's response box. Feedback was provided after each response. The signal level was varied adaptively with a stepsize of 2 or 1 dB. Initially, the signal level was set well above masked threshold, and the level was reduced by 2 dB after three successive correct responses; after the first incorrect response the level was increased by 2 dB and the stepsize was reduced to 1 dB. A three-down one-up procedure (Levitt, 1971) was applied in order to estimate the level that would produce 79.4 %

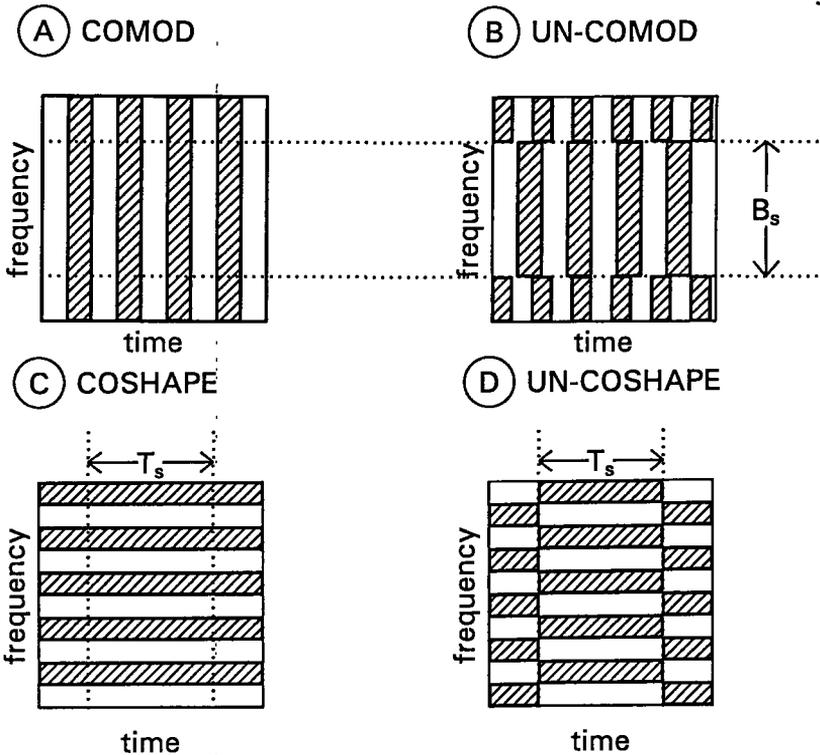


Fig. 1 Illustration of the four specific masker modulation patterns. The patterns represent temporally modulated patterns (A and B) and spectrally shaped patterns (C and D). The bandwidth and duration of the signal are indicated by B_s and T_s , respectively. Patterns C and D are obtained from patterns A and B, respectively, by interchanging the time and frequency axis.

correct responses. After three reversals the data collection started up to a total of thirteen reversals. Threshold was defined as the mean of the signal levels at the last ten reversals in either direction. One series, with three adaptive runs for the same signal, took about 10 minutes. At least six such adaptive runs were obtained for each stimulus condition.

3.2.2 Subjects

Three subjects participated in the experiments. One subject (the first author) was experienced in psychoacoustical experiments, the other two were students, who were paid an hourly wage. The students were each given two hours of daily practice for a period of one week after which their thresholds had stabilized. After this training period the data collection started.

3.2.3 Apparatus

The experiments were performed under control of an IBM-AT computer. Signals were generated, at 40- μ s sample intervals, using a digital signal processor based on a TMS32010. Broad-band noise was generated, using a Wandel & Golterman RG-1 analog noise generator. The 100% SAM noise was created by leading the noise from the noise generator through an analog amplitude modulator (TNO, custom built) of which the modulation depth and modulation frequency could be programmed. For the comodulated noise (COMOD) the modulation frequency was set at 32 Hz. The uncomodulated noise (UN-COMOD) was created by leading noise from the same noise generator to a second analog amplitude modulator operating at a modulation frequency of 40 Hz. The on-signal masker band was modulated at 32 Hz, the off-signal masker at 40 Hz. The output of the first modulator was band-pass filtered, the output of the second one band-reject filtered, in each case by two serially connected Krohn-Hite 3341 filters.

Noise with a cosine-shaped power spectrum was created by sampling the noise with a 16-bit Analog-to-Digital converter (TNO, custom built) at a sample frequency of 25 kHz and adding it to a delayed version of the same noise. The delay and polarity determined the spacing and the positions of the peaks in the cosine-rippled power spectrum. The inter-peak-spacing was set at 800 Hz, with a peak or a valley at 1.6 kHz. For the coshaped noise condition (COSHAPE), the spectral shaping was identical for the on-signal interval, T_s , and the flanker intervals preceding and following the on-signal interval (see also Fig.1). The un-coshaped condition (UN-COSHAPE), with different spectral shaping of the on-signal masker and the flanker bursts was created by inverting the spectrum of the on-signal masker in the two flanker bursts. The total stimulus was bandpass filtered from 360 up to 7180 Hz by two serially connected Krohn-Hite 3341 filters.

The stimuli were presented diotically (same stimulus to both ears) by means of Beyer DT-48 headphones. Subjects were seated in a soundproof room.

3.2.4 Level definition

The energy spectra of all stimuli were determined using a real-time frequency analyzer (Brüel & Kjær 2123). All maskers were derived from white noise. The noise level was defined by the spectral density N_0 , the mean noise intensity in a 1-Hz interval. For temporally modulated noise, this N_0 represents the intensity averaged over a modulation cycle. For spectrally

shaped noise, a similar definition is applied: N_0 is the average intensity over one spectral modulation cycle, being 3 dB below the peaks of the cosine-shaped power spectrum. In all experiments the level of the masking noise was set to approximately 50 dB SL, corresponding to a noise spectral density, in case of unmodulated noise, of about 20 dB/Hz. The threshold levels of the signals were defined in terms of their total energy normalized by the spectral density of the masking noise: E_{tot}/N_0 (van den Brink and Houtgast, 1990).

Results averaged across trials per condition and subjects (typically 6x3 values) are presented. The error bars in the figures represent the standard deviations. The main variance in the data resulted from a change in stimulus conditions; interactions between subjects and conditions were insignificant.

3.3 EXPERIMENT I: NOISE BANDWIDTH

Rather than using the standard CMR stimulus configuration, with multiple comodulated narrow-band maskers, a broad-band white-noise masker was used, 100% sinusoidally amplitude-modulated at 32 Hz. In this first experiment, using both long- and short-duration narrow-band signals, we investigated whether this approach yields CMR similar to the CMR effect reported by Hall et al. (1984).

3.3.1 Stimuli

The short-duration signal was an impulse response of a filter with a bandwidth of 1/3-oct centered at 1.6 kHz. The long-duration signal was a 1.6 kHz sinusoid, with a steady state duration of 75 ms, and 5 ms cosinusoidal rise-fall times. The masked thresholds of these signals were determined for masker bandwidths of 1/3, 3/3, 5/3, 9/3 or 13/3 octs centered around 1.6 kHz. The unmodulated noise condition is referred to as REF, and the comodulated noise masker condition (modulation frequency of 32 Hz) as COMOD. With the short-duration signal the masked thresholds were determined for the REF masking condition and for the COMOD masking condition, for both a peak and a valley placement of the signal. For the long-duration signal the masked thresholds were determined for both the REF and the COMOD masking condition. The long-duration signal was presented at random relative to the modulation cycle of the masker.

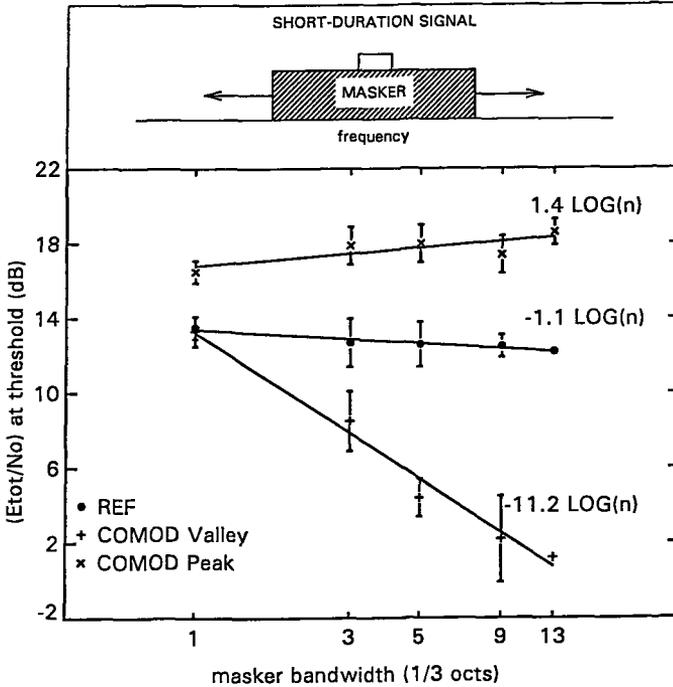


Fig. 2 The stimulus condition is illustrated in the upper panel. The lower panel shows the masked thresholds of a short-duration narrow-band (1/3-oct) impulse response as a function of masker bandwidth. The masker was either unmodulated white noise (REF) or 100% SAM white noise (modulation frequency 32 Hz), with the signal in the valley or at the peak of the temporal envelope of the masker (COMOD Valley and COMOD Peak, respectively). The error bars indicate the standard deviations for 3 subjects and 6 trials per subject.

3.3.2 Results and discussion

Figs. 2 and 3 show the masked thresholds as a function of masker bandwidth, for the short- and long-duration signal, respectively.

The datapoints for each masking condition have been subjected to a least-squares linear regression fit ($y = \alpha + \beta \log(n)$, with n being the number of 1/3-oct bands involved). The β -values, reflecting the effect of masker bandwidth on the masked threshold, are presented in the graph.

The data for the short-duration signal, as presented in Fig. 2, show that the masked thresholds in the REF masking condition are largely independent of masker bandwidth. The data for the COMOD masking condition in Fig. 2 reveal CMR (a decrease in masked threshold with increasing masker bandwidth), but only when the signal is placed in the valley of the masker

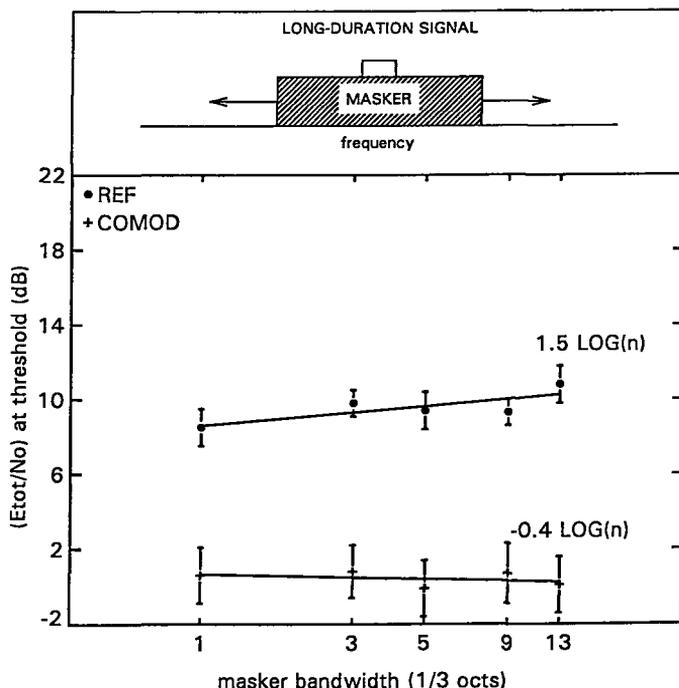


Fig. 3 The stimulus condition is illustrated in the upper panel. The lower panel shows the masked thresholds of a long-duration (75-ms) pure-tone at 1.6 kHz as a function of masker bandwidth. The masker was either unmodulated white noise (REF) or 100% SAM white noise (COMOD, modulation frequency 32 Hz). The error bars indicate standard deviations.

envelope. No CMR is found when the signal is placed at the peak of the masker envelope.

For the long-duration signal the data for both masking conditions, as presented in Fig. 3, are about independent of the noise bandwidth.

In contrast to data collected by Hall et al. (1984), our data for the long-duration signal indicate that adding 1/3-oct COMOD noise bands does not facilitate detection. The data do show a release from masking with modulation of about 10 dB (the difference between REF and COMOD) starting already at a masker bandwidth of 1/3-oct, which suggests a within-channel detection cue.

It is most remarkable that the masked threshold for the COMOD masking condition with the signal placed in a valley of the SAM noise is equal to the masked threshold of the short-duration signal presented in unmodulated noise (REF) of 1/3-oct bandwidth. This indicates that in the first condition, the auditory system does not profit at all from the low noise level in the valleys of the modulated noise, although all three subjects reported that the modulation

pattern was clearly audible. The audibility is in line with modulation detection data of Viemeister (1979). Similar results have been reported by Hall et al. (1984), showing that the masked thresholds for a short-duration signal in the REF and COMOD conditions at 1/3-oct noise bandwidth are about the same. This has been interpreted as being related to the amount of temporal uncertainty for such a brief stimulus.

Thus, a comodulation release from masking is found for a short-duration signal presented at a temporal valley of the masker envelope. For this signal the CMR increases with increasing noise bandwidth in accordance with a decrease of $11.2 \log(n)$ in masked threshold, where n represents the number of comodulated 1/3-oct noise bands. For a long duration signal the data suggest no CMR.

3.4 ACROSS-FREQUENCY INTEGRATION OF SIGNAL ENERGY AND CMR

3.4.1 Control experiment

When increasing signal bandwidth, the spectral levels should be chosen such that every individual 1/3-oct signal component contributes equally to the masked threshold in order to avoid that one channel solely determines detection. For unmodulated white noise, it was determined earlier (van den Brink and Houtgast, 1990) that equal $(E/N_0)_{1/3}$ (E/N_0 of the individual 1/3-oct bands excited by the signal should be equal to one another) could be taken as a first order approximation for equal-detectability of the individual components within a broadband signal. A control experiment was performed to investigate the applicability of this rule in case of a temporally modulated masker. The masked thresholds of four 1/3-oct bandpass impulse responses at center frequencies of 0.5, 1.0, 2.0 and 4.0 kHz were determined, for a signal placement in the valley and at the peak of the envelope of a 32 Hz 100 % sinusoidally amplitude-modulated noise. To minimize the effect of CMR on the masked thresholds of the four signals, the masker was restricted to a 1/3-oct centered at the signal frequency (in the previous paragraph no difference in masked thresholds between the COMOD and REF condition was found for this bandwidth).

For all three subjects the masked thresholds at these four frequencies were, both in case of a valley or a peak placement, determined by about the same amount of energy. The results were subjected to a linear regression fit yielding slopes of 1.2 dB/oct and 1.3 dB/oct as a function of frequency, for a

valley and a peak placement, respectively. The mean threshold values and standard deviations are 13.9 dB and 1.7 dB, and 17.7 dB and 2.3 dB, for a valley and a peak placement, respectively. This indicates that, also in case of modulated maskers, equal $(E/N_0)_{1/3}$ can be used as a first order measure for equal-detectability over the frequency scale for broadband signals.

3.4.2. Short-duration signals

1. Stimuli

A set of five impulse responses of filters was constructed with a geometrical center frequency of 1.6 kHz and bandwidths of 1/3, 3/3, 5/3, 9/3 and 13/3 octs, respectively. The masked thresholds of these signals were determined for two types of masking noise: unmodulated (REF condition) and 100% SAM noise (COMOD condition). The masker bandwidth was always set to 13/3-octs centered at 1.6 kHz. For the COMOD masking condition masked thresholds were determined for signal placements at the peak and in the valley of the masker envelope. The modulation frequency was set at 32 Hz.

2. Results and discussion

The results are summarized in Fig. 4. Again, the datapoints for each masking condition were subjected to a least-squares linear regression fit ($y = \alpha + \beta \log(n)$, n being the number of 1/3-oct signal components) and the β -values are included in Fig. 4.

The data for the REF masking condition follow $3.2\log(n)$ which is in line with earlier results on across-frequency integration effects on the masked threshold (van den Brink and Houtgast, 1990).

For the peak placement in the COMOD masking condition the slope ($1.9\log(n)$) is not very different. The average difference between these data and those found for the REF conditions amounts to about 6 dB. This is slightly more than expected when the masked thresholds would be determined by the instantaneous signal-to-noise ratio (For SAM noise, the difference between peak and rms intensity is 4.2 dB). This indicates that additional masking occurs. When the signal is placed in the valley of the COMOD masker and signal bandwidth is increased we also find a slope ($4.7\log(n)$) close to the REF condition. According to the explanation of CMR given by Cohen and Schubert (1985,1987) and Richards (1987) one might expect a larger effect of signal bandwidth because at the largest bandwidth of 13/3 octs the signal bandwidth matches the masker bandwidth. The explanation given by Cohen and Schubert (1985,1987) and Richards (1987) would suggest that the masked threshold at this condition equals that found for a 1/3-oct signal and

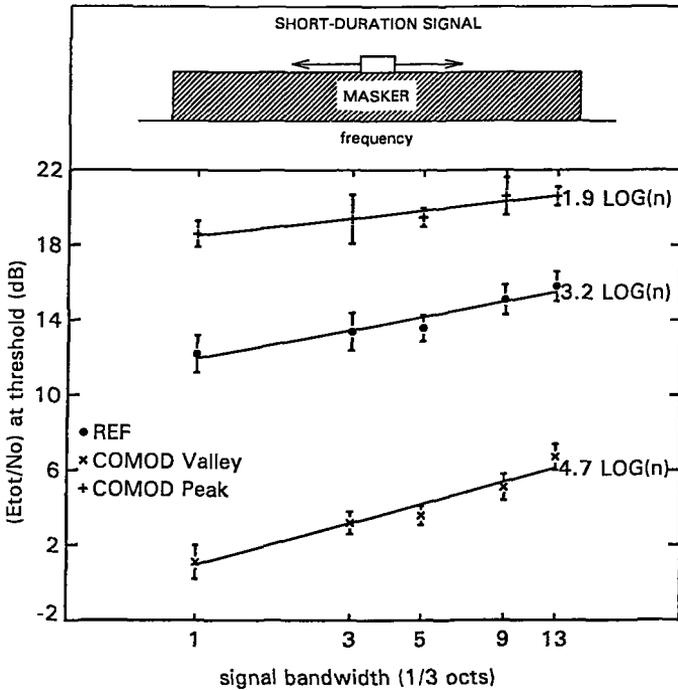


Fig. 4 The stimulus condition is illustrated in the upper panel. The lower panel shows the masked thresholds of a short-duration bandpass impulse response as a function of signal bandwidth. The masker, having a bandwidth of 13/3 octs, was either unmodulated white noise (REF) or 100% SAM white noise (modulation frequency 32 Hz), with the signal in the valley (COMOD Valley) or at the peak (COMOD Peak) of the masker envelope. The error bars indicate standard deviations.

masker bandwidth which is about 14 dB (Fig. 2). The present result suggests that on-signal co-modulated masker bands contribute considerably to a release from masking. This has already been mentioned by Hall et al. (1988), who found that CMR occurs when each comodulated noise band contains a pure-tone signal.

3.4.3 Long-duration signals

1. Stimuli

Five signals were constructed, all with a geometrical center frequency of 1.6 kHz, one pure-tone signal and four complex signals with 3, 5, 9 and 13 components at 1/3 oct. distances, respectively. The duration of the signals was 75 ms with 5-ms cosinusoidal rise-fall times. The masked thresholds of these signals were determined for two types of masking noise: unmodulated (REF

condition) and 100% sinusoidally amplitude-modulated noise (COMOD condition). The masker bandwidth was set to 13/3-octs centered at 1.6 kHz. For the COMOD masking condition the modulation frequency was set at 32 Hz

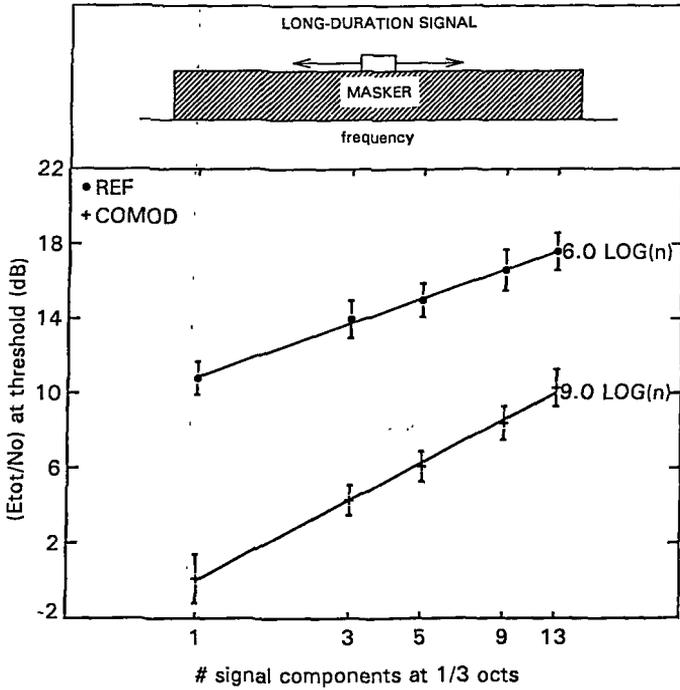


Fig. 5 The stimulus condition is illustrated in the upper panel. The lower panel shows the masked thresholds of long-duration signal as a function of the number of components at 1/3-oct intervals. The masker, having a bandwidth of 13/3 octs, was either unmodulated white noise (REF) or 100% SAM white noise (COMOD, modulation frequency 32 Hz). The error bars indicate standard deviations.

2. Results and discussion

The results are presented in Fig. 5, which shows the masked thresholds of the long-duration signals, as a function of signal bandwidth, for an unmodulated masker (REF condition) and a comodulated masker (COMOD condition). The datapoints for each masking condition were subjected to a least-squares linear regression fit ($y = \alpha + \beta \log(n)$, n being the number of signal components at 1/3-oct intervals), and the β -values were included in Fig. 5.

The result of $6\log(n)$ for the REF condition is in line with earlier results on across-frequency integration for long-duration signals (e.g. Buus et al., 1986 and van den Brink and Houtgast, 1990). The result of $9\log(n)$ for the

COMOD condition indicates almost none across-frequency signal integration ($10\log(n)$ implies no integration).

3.4.4 Uncomodulated masking condition (UN-COMOD)

The most proper measure of the effect of co-modulated flankerbands is obtained when the thresholds found for the comodulated flankerbands are compared to those found for uncomodulated flankerbands (CMR(U-C)). This is in line with the literature (e.g. Hall et al., 1984, Moore ,1990). In this section we report the effect of signal bandwidth on this measure. The noise bands in which no signal components are presented were modulated at 40 Hz, and the noise bands containing signal components were modulated at 32 Hz. Thus, when increasing signal bandwidth the number of comodulated masker bands is increased while the number of uncomodulated flankerbands is decreased (lower and upper scale of Fig. 6, respectively). For this condition we measured the masked thresholds for both the long-duration signals and short-duration signals (peak and valley placement). The results are summarized in Fig. 6.

The zero value of CMR(U-C) at the signal bandwidth of 13/3 octs is trivial. In the absence of flankerbands the COMOD and the UN-COMOD masking conditions are identical. In agreement with the results from Fig. 2 and Fig. 3 the data in Fig. 6 show that the CMR(U-C) is much larger for short-duration signals placed at a temporal valley of the masker envelope than for the long-duration signal. However, for the short-duration signal, Fig. 2 showed a difference between the thresholds found for the 13/3-oct and the 1/3-oct masker bandwidth of about 12 dB, and, for the long-duration signal, Fig. 3 showed a difference of about 0 dB. CMR(U-C) in Fig. 6 equals 16 dB and 4 dB, respectively. Adding the uncomodulated flankerbands introduces an increase of about 4 dB in masked threshold with respect to the conditions in Fig. 2 and Fig. 3 without flankerbands. This suggests across-frequency integration of masker energy, also called across-channel masking (ACM, Moore 1990). When the short-duration signal is placed at a temporal peak of the masker envelope essentially no CMR(U-C) effect is observed.

The data in Fig. 6 indicate that a substantial CMR-effect is present due to flankerband comodulation, which decreases with a decreasing number of flankerbands, while signal bandwidth increases.

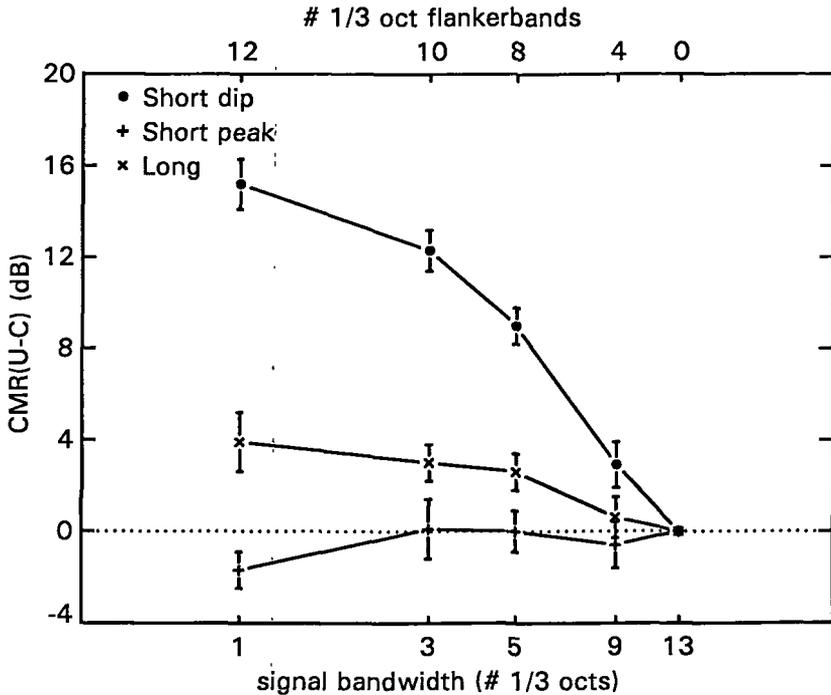


Fig. 6 CMR(U-C) as a function of increasing signal bandwidth (or decreasing number of 1/3-oct flankerbands) for a long-duration signal and a short-duration signal placed in the valley and at the peak of the masker envelope. The total masker bandwidth was always 13/3 octs. The masker bands containing signal components were modulated at a modulation frequency of 32 Hz whereas the flanking masker bands were modulated at a modulation frequency of 40 Hz. The error bars indicate standard deviations.

3.5 TEMPORAL RESOLUTION AND THE WIDTH OF THE EAR'S TEMPORAL WINDOW

An interesting aspect of the previous experiment concerning the influence of masker bandwidth on the masked threshold of a 1/3-oct short-duration signal, is the fact that the data in Fig. 2 show that the peak-to-valley difference depends on masker bandwidth. Green (1973) reported a series of experiments performed by Viemeister who used the difference between peak and valley masked thresholds to determine the integration time constant of the auditory system. The data in Fig. 2 suggest that this integration time depends on masker bandwidth. An experiment was performed to illustrate the influence of masker bandwidth on the integration time of the auditory system, as estimated from the peak-to-valley difference.

3.5.1 Stimuli

The signal consisted of an impulse response of a filter centered at 1.6 kHz with a bandwidth of 1/3-oct. The masked thresholds of this signal were determined for a COMOD masking condition for both a valley and a peak placement of the signal. The bandwidth of the 100% SAM noise was set to either 1/3-oct or 3-octs. The modulation frequency of the 100% SAM noise was set to 8, 16, 32, 64 or 128 Hz.

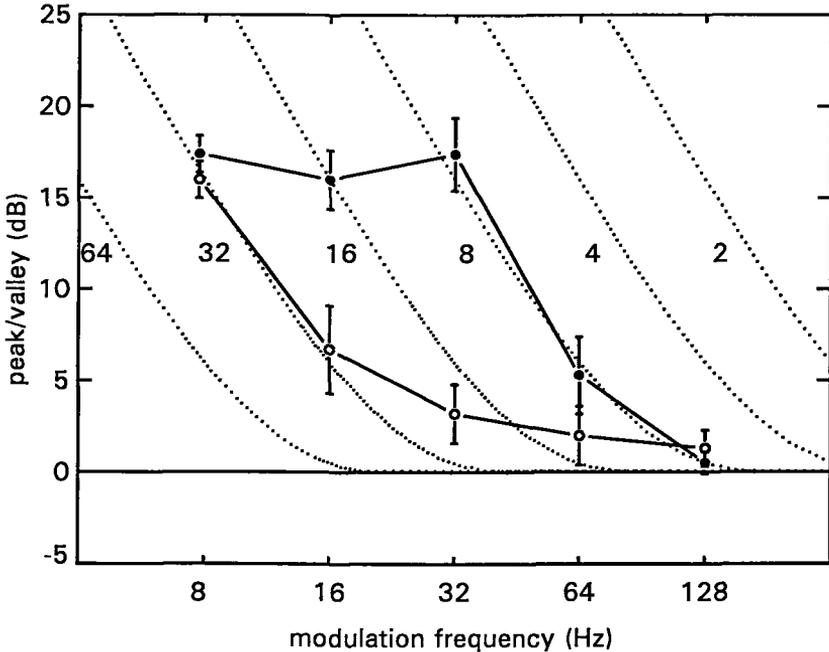


Fig. 7 Peak-to-valley distance in dB as a function of the temporal modulation frequency of the masker, for SAM noise maskers at a bandwidth of either 1/3-oct (open circles) or 13/3-oct (filled circles). The signal was an impulse response of a filter with a bandwidth of 1/3-oct centered at 1.6 kHz. In order to provide an global estimate of the ear's temporal window the theoretical peak-to-valley distances as a function of the modulation frequency have been included (dotted curves). These theoretical curves are based on a Gaussian temporal window with an effective width in ms as indicated. The error bars indicate standard deviations.

3.5.2 Results and discussion

The results are plotted in Fig. 7, which shows the difference between the threshold levels measured at the peak and in the valley, for noise bandwidths of 1/3-oct and 3-octs, as a function of the modulation frequency of the masker. Fig. 7 also shows the theoretical peak-to-valley differences as a function of the modulation frequency, when the 100% SAM noise is subjected

to an intensity-weighting Gaussian temporal window, for several widths of this Gaussian window (See Appendix). Although the Gaussian-shaped temporal window only provides a first-order approximation of the ear's temporal window, it can be applied to interpret the experimental results quantitatively. The interesting part of the curves, at low peak-to-valley differences, is fitted reasonably well by the theoretical curves. We did not attempt to model the upper border at the high peak-to-valley differences which is related to the absolute threshold. The data for the wide-band masker are in line with data reported by Moore et al. (1988), yielding a width of the temporal window of about 8 ms. However, the data for the 1/3-oct masker bandwidth give an estimated width of the ear's temporal window of about 32 ms.

3.6 THE SPECTRAL ANALOG OF CMR

In this second part of the paper we will present some experiments focussed on the question of whether a release from masking occurs for the detection of a tone pulse masked by simultaneously presented noise (masker burst) with a spectrally shaped profile, when a masker with an identical profile precedes and follows the signal (temporal flanker bursts). If such a phenomenon occurs, we would call it, in parallel with CMR coshaping release from masking (CSMR). As a spectral equivalent of the temporal sinusoidal amplitude-modulation, the power spectrum was cosine-shaped, thus being sinusoidally intensity-modulated. The difference between amplitude and intensity modulation is thought to be of minor interest. For the coshaping condition, the spectral shape of the noise does not change during the total masker duration; the spectral shape of the flanker bursts is identical to that of the masker burst (Fig. 1c) and they follow each other without intermission. For the un-coshaped masking condition two spectral shapes were used: one for the noise presented simultaneously with the signal (signal burst) and another one for the noise presented nonsimultaneously with the signal (flanker burst). The spectral peak-to-peak distance was identical for these two shapes, but the peaks and valleys were interchanged (Fig. 1d).

The experiments are related to experiments as performed by Carlyon (1989). He reported a series of experiments in which thresholds were measured for 5-ms 1-kHz tones masked by a synchronous (5 ms) noise burst containing a spectral notch at the signal frequency. Thresholds were reduced by prior exposure to a noise burst that had the same spectral shape as the on-signal masker suggesting a possible CSMR effect.

In the present series of experiments we will investigate CSMR in a CMR-like experimental approach.

3.6.1 Masker duration

1. Stimuli

In this experiment a 1.6 kHz pure tone pulse, with 5-ms cosinusoidal rise-fall time and a steady-state duration of 18.75 ms, was used as a signal. Two types of noises were used: white noise (REF) and coshaped noise (COSHAPe) with a peak-spacing of 800 Hz, and either a peak or a valley in the spectrum at the signal frequency (1.6 kHz). The masked threshold of the tone pulse in the two masker conditions was determined as a function of masker-plus-flanker duration, which was set to either 18.75, 37.5, 75, 150 or 300 ms, with 5-ms cosinusoidal rise-fall times. Masker duration was always 18.75 ms. Thus, with increasing total duration the duration of the flanker bursts was increased. The signal was presented at the temporal center of the masker.

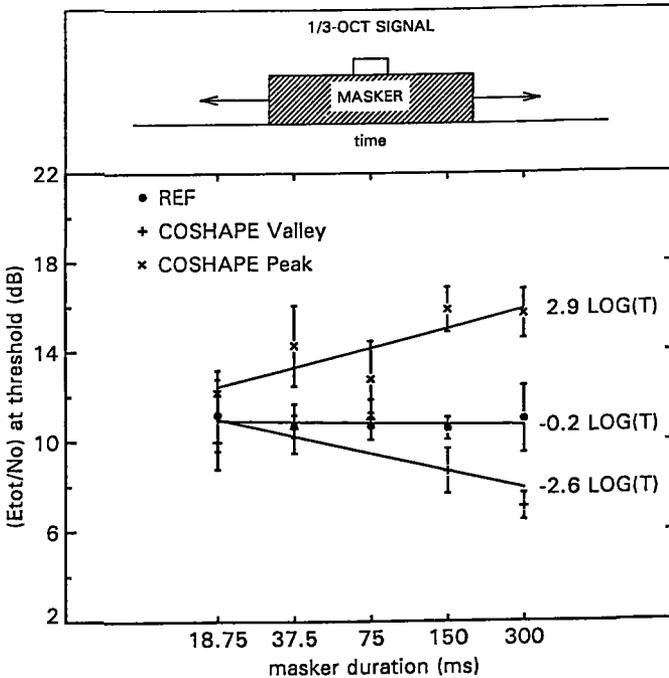


Fig. 8 The stimulus condition is illustrated in the upper panel. The lower panel shows masked thresholds of a 18.75 ms 1.6 kHz pure-tone, as a function of masker duration. The masker was either white noise (REF) or noise with a cosine peak-rippled power spectrum (COSHAPe), with an inter-peak spacing of 800 Hz and with either a peak or a valley at the signal frequency. The error bars indicate standard deviations.

2. Results and discussion.

The results are plotted in Fig. 8, which shows the masked threshold of the tone pulse for the three masking conditions as a function of total noise duration.

The data for each of the masker conditions were subjected to a linear regression fit ($y = \alpha + \beta \log(T/T_s)$, $T_s = 18.75$ ms), yielding β -values of -0.2, -2.6 and 2.9 and α -values of 10.9, 11.0 and 12.5, for the REF masking condition and the COSHAPE masking condition with the signal in a spectral valley or at a spectral peak, respectively.

For the REF condition the masked threshold of the tone pulse remains the same for all masker durations. This reflects the fact that the signal duration of 18.75 ms is larger than the 'critical masking interval' as determined by Penner and Cudahy (1973). For a masker duration equal to the signal duration the masked thresholds for the three masking conditions are about equal to one another, indicating that the spectral modulation does not affect signal detection, although all subjects reported that they were able to distinguish the different noise types.

When increasing masker duration the masked threshold of the tone pulse placed at a spectral peak increases with $2.9\log(T/T_s)$, whereas for a valley placement the masked threshold decreases with $2.6\log(T/T_s)$. The data suggest that some coshaping release (CSMR) from masking occurs only in the valley, whereas a reverse effect occurs at the peak.

3.6.2 Coshaped versus Un-coshaped masking condition

In order to determine whether the CSMR, as presented in the previous paragraph, is a consequence of the coherence (in time) of the spectral envelope an experiment on the difference between coshaped and un-coshaped maskers was performed as a function of signal duration.

1. Stimuli

Analogous to the UN-COMOD experiment reported in Sec. III D., we studied the effect of un-coshaped flanker bursts as a function of the duration of the signal. A set of five pure-tone pulses at 1.6 kHz, with steady-state durations of 18.75, 37.5, 75, 150 and 300 ms and 5-ms cosinusoidal rise-fall times was used. The total masker-plus-flanker duration was set to 300 ms. Two types of masking noise were used, coshaped noise and un-coshaped noise. The signals were presented at the temporal center of the masker. The masked thresholds of the signals were determined for both a peak and a valley placement at the spectral envelope. The peak-spacing of the masker was set to

800 Hz. The un-coshaped flanker burst was simply the spectrally inverted on-signal masker with the same peak-spacing (Fig. 1d).

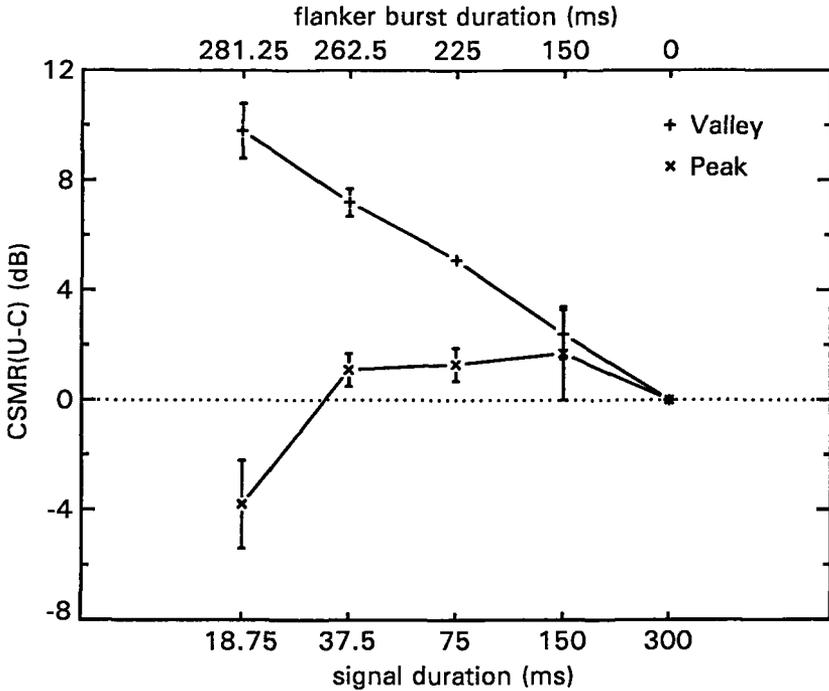


Fig. 9 CSMR(U-C) as a function of signal duration in a 300-ms masker for a pure-tone pulse placed in the spectral valley or at the spectral peak of the masker spectral envelope. The error bars indicate standard deviations.

2. Results and discussion

The results are plotted in Fig. 9, which shows the difference in masked thresholds as measured with the un-coshaped and coshaped masking condition as a function of signal duration. This difference is expressed as CSMR(U-C).

Similarly to the result in Fig. 6 the zero value of CSMR(U-C) at 300 ms signal duration is trivial. When the signal duration equals the total noise duration there are no flankerbursts and the coshaped and un-coshaped masking conditions are simply identical.

It can be seen from Fig. 9 that a substantial CSMR(U-C) is found for a valley placement of the signal. Essentially no CSMR is found when the signal is placed at a peak of the spectral envelope of the masker, except for the 18.75 ms signal duration where a 4-dB negative CSMR(U-C) is found. In the

latter condition we found a decrease of the masked threshold for the UN-COSHAPE condition which may be related to the clear temporal contrast marking, in this condition, the interval in which the signal is to be expected.

3.7 SPECTRAL RESOLUTION AND THE AUDITORY BANDWIDTH

The peak-to-valley ratio has been used to determine the auditory spectral resolution (Festen, 1983), and to estimate the auditory bandwidth (Houtgast, 1974). However, the data in Fig. 8 indicate that the peak-to-valley distance estimated from threshold measurements depends on masker duration. Therefore we performed an experiment to determine the effect of masker duration on the peak-to-valley distance as a function of the peak spacing (ripple density), in order to estimate the auditory bandwidth for two different masker durations.

3.7.1 Stimuli

The same 1.6 kHz tone pulse with a steady-state duration of 18.75 ms as in the previous experiment was used. The masked thresholds were determined for a COSHAPE masking condition. The peak-spacing of the coshaped noise was used as a parameter; the duration of the noise masker was fixed to either 18.75 ms or 300 ms. The peak-spacing of the masker was set to 200, 400, 533, 800 or 1600 Hz, respectively. The total stimulus was bandpass filtered within a spectral range from 360 to 7180 Hz.

3.7.2 Results and discussion

The results are plotted in Fig. 10, which shows the peak-to-valley threshold difference for the 18.75 ms tone pulse in either a 18.75 ms COSHAPE masking condition or a 300 ms COSHAPE masking condition, as a function of the ripple density of the masker. In addition, the theoretical peak-to-valley distances have been plotted, when applying the peak-rippled noise to a Gaussian intensity-weighting function with bandwidth B (See Appendix). Although the Gaussian window only approximates the shape of the ear's window the curves can be used to interpret the experimental data.

The datapoints for the 300 ms masker indicate that the auditory bandwidth is about 250 Hz, whereas the datapoints for the 18.75 ms masker suggest an auditory bandwidth of about 500 Hz at 1.6 kHz. This indicates, that the spectral resolution and subsequently the estimated auditory bandwidth,

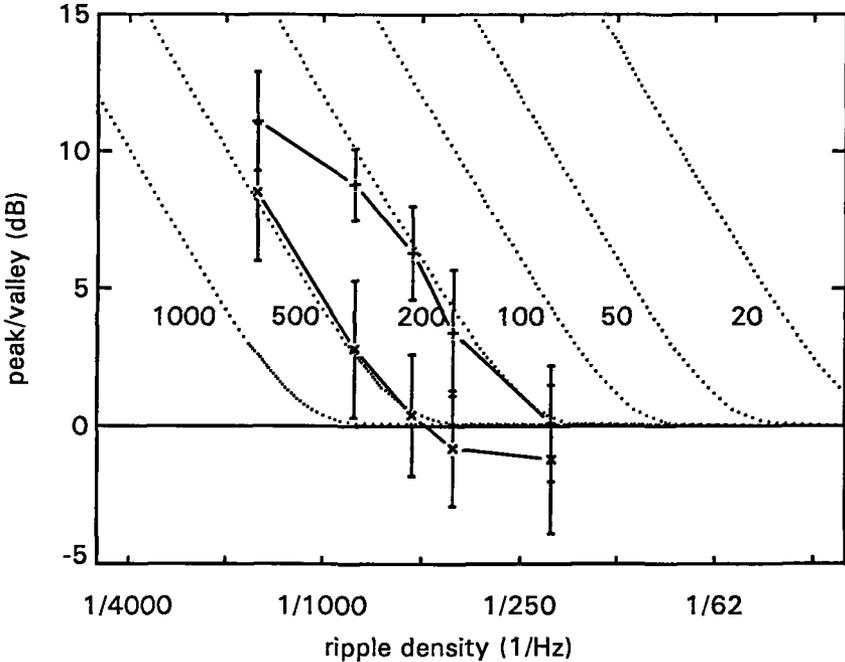


Fig. 10 Peak-to-valley distance in dB as a function of the ripple density of the sinusoidally rippled power spectrum noise. These peak-to-valley distances are presented for two masker durations, 18.75 ms (cross symbols) and 300 ms (plus symbols), while the tone duration was 18.75 ms. The theoretical peak-to-valley distances as a function of the ripple density have been included (dotted curves) in order to provide an global estimate of the auditory bandwidth. These theoretical curves are based on a Gaussian spectral window with an effective width in Hz as indicated. The error bars indicate the standard deviations.

as determined from the peak-to-valley distance, depends on masker duration. The data for the 300 ms duration masker are in agreement with critical bandwidth data from Scharf (1970).

3.8 GENERAL DISCUSSION

The experiments presented in the first part of this paper reveal that a CMR effect, as traditionally obtained with multiple narrow band comodulated maskers at some distance from the on-signal masker, can also be obtained with contiguous on-signal maskers and flankerbands. Substantial CMR (a decrease of masked threshold with increasing masker bandwidth) is found only for a short-duration narrowband signal placed at a valley of the masker envelope.

A most striking result is that the masked thresholds of a narrowband brief signal presented in either 1/3-oct unmodulated white noise or in 1/3-oct

100% SAM white noise at the peak or in the valley are about equal to one another (Fig. 2). Preliminary detection models showed that the decrease in masked threshold with increasing masker bandwidth of a signal placed in the valley of 100% SAM noise could be related to decreasing temporal uncertainty across masker and flanker bands.

In contrast to results presented in the literature (Hall et al., 1984) no release from masking was found for long-duration signals when increasing masker bandwidth (Fig. 3). It might be argued that the discrepancy between the present results for long-duration signals and those of Hall et al. (1984) is a consequence of the smaller bandwidths ($< 1/3$ -oct) in Hall's study.

Considering the effect of noise bandwidth for the 100% SAM white noise (Figs. 2 and 3), an interpretation in terms of nonsimultaneous masking instead of comodulation release from simultaneous masking seems feasible. In experiments on non-simultaneous masking (e.g. forward masking), the masked threshold is found to decrease with masker bandwidth, which has been ascribed to the effect of lateral suppression (Houtgast, 1974). In the present study, detection of a signal in the valley of 32-Hz 100% SAM noise could actually constitute a condition of nonsimultaneous masking, in that the threshold is determined by forward masking produced by the peak preceding the valley. However, this would apply not only to the short-duration signal in the valley, but also to the long-duration signal running through several valleys of the masker. Since no release of masking with increasing bandwidth was found for long-duration narrowband signals (Fig. 3), such an interpretation in terms of lateral suppression revealed in nonsimultaneous masking does not seem applicable.

The experiments presented in the first part of this paper reveal a new aspect of the CMR phenomenon. While all current models assume that CMR is based on an across-frequency channel envelope disparity, our data show that substantial CMR remains present when increasing signal bandwidth in the presence of a wide-band masker (Figs. 4 and 5). The increasing signal bandwidth implies a diminishing of the amount of across-channel envelope disparity. Co-modulation masking release also occurs for on-signal masker bands.

In the second part of the paper it was shown that, similarly to comodulation release from masking (CMR) in temporally shaped noise, coshaping release from masking (CSMR) in spectrally shaped noise occurred. A release from masking was found for the detection of a tone pulse masked by simultaneously presented noise with a spectrally shaped profile, when additional noise, temporal flanker burst, with identical spectrally-shaped

profiles were added preceding and following the signal. As with CMR, this release from masking was found only when the test tone was placed in the valley of the spectral masker profile. For a peak placement of the test tone, no CSMR was found. This implies that, with CMR and CSMR, the difference between the masked thresholds for a peak and a valley placement of the test tone increases. This has interesting implications for the determination of auditory spectral and temporal resolution. When estimating auditory resolution from the peak-to-valley differences in masked threshold measured for temporally modulated or spectrally shaped noise, it was found that temporal resolution depends on masker bandwidth, and that spectral resolution depends on masker duration.

Festen (1983) performed an experiment in order to determine the ear's temporal resolution. He argued that the differences in integration time, estimated from peak-to-valley differences, as reported in various studies, depend not only on the cutoff frequency of the low-pass filter but also on the shape of the filter, and that other differences must have been caused by differences in click frequency and the use of amplitude-modulated rather than intensity-modulated noise. However, an important additional reason for these differences is illustrated in Fig. 7; the time constant as derived from the peak-to-valley difference is highly dependent on masker bandwidth. Also, when using the peak-to-valley difference in masked threshold for spectrally shaped noise, many different values can be found for the critical bandwidth since the bandwidth determined in this way depends on the duration of the masker.

3.9 CONCLUSIONS.

1. CMR is found for a short-duration signal placed in the valley of the temporal envelope of a 100% SAM broadband masker, similar to traditionally reported CMR when using multiple comodulated narrow bands of noise separated from the on-signal masker (Fig. 2).
2. When increasing signal bandwidth of a short-duration signal placed at a valley of a 100% SAM wide-band noise, the masked threshold only increases slightly, indicating that comodulated masker bands containing a signal component contribute to CMR (Fig. 4).
3. Whereas all current theories assume that CMR is based on across-channel envelope disparity, our data suggest that substantial CMR remains when increasing signal bandwidth in the presence of a wide-band masker. The increasing signal bandwidth implies a diminishing amount of across-channel envelope disparity.

4. Similar to CMR a coshaping release from masking (CSMR) has been found for signal detection in spectrally shaped noise. This CSMR is found only for a placement of the signal in the valley of the spectral envelope of the masker.
5. The decrease in the masked threshold of a narrowband brief signal when placed in the valley of a masker depends on masker duration (for a valley in the frequency domain) or on masker bandwidth (for a valley in the time domain).

ACKNOWLEDGEMENT

This investigation was supported by the Netherlands Organization for Scientific Research.

APPENDIX

A Gaussian intensity-weighting function with effective duration T (Equivalent Rectangular Duration = ERD) is given by the following expression:

$$W(t) = \exp\left(-\pi \frac{[t-t_0]^2}{T^2}\right) . \quad (\text{A1})$$

For 100% SAM noise, the intensity function is:

$$I(t) = \left[1 \pm \cos\left(\frac{2\pi (t-t_0)}{\Delta t}\right) \right]^2 , \text{ with } \Delta t = \frac{1}{f_{\text{mod}}} . \quad (\text{A2})$$

The plus sign refers to the condition "filter centered at top" and the minus sign to "filter centered at valley". The weighting-function output, for the peak or valley condition, is determined by the following integral, with the plus sign for "filter at top" and the minus sign for "filter at valley":

$$\begin{aligned} \text{Output}(\pm) &= \int W(t)I(t) dt = \\ &= 1.5 T \pm 2 T \exp\left(\frac{-\pi T^2}{\Delta t^2}\right) + 0.5 T \exp\left(\frac{-4\pi T^2}{\Delta t^2}\right) . \quad (\text{A3}) \end{aligned}$$

The peak-to-valley distance in dB is obtained from $10 \log(\text{output}(+)/\text{output}())$. Using a similar procedure with a Gaussian intensity weighing function in the frequency domain (Equivalent rectangular bandwidth = B) applied to the peak-rippled noise (SIM) the outputs for the peak and the valley conditions are:

$$B \pm B \exp\left(-\frac{\pi B^2}{\Delta f^2}\right) . \quad (\text{A4})$$

SUMMARY

This summary presents a general synopsis of the results of this thesis. Conclusions are provided in each individual chapter.

The main issue of this thesis is the effect of signal pattern on signal detection in noise. In this context signal pattern is the distribution of signal energy over frequency and time. With respect to the distribution over frequency, it was important that the spectral shapes of the signal and the noise ensured a condition of equal-detectability for each individual frequency channel so that detection was not simply determined by the channel with the highest signal-to-noise ratio (Chapter 1). (Throughout this study, an auditory channel, or critical band, is supposed to have a width of 1/3-octave band.) This condition is approximated by using conditions with equal $(E/N_0)_{1/3}$ (E specifying the energy of a signal within a 1/3-oct band and N_0 specifying the local noise spectral density in that band). Thus, in case of a white noise masker (N_0 constant over frequency), the energy of the individual 1/3-oct signal components should be equal in order to ensure the condition of equal-detectability within all individual channels covered by the signal.

Given this constraint, the effect of signal pattern on signal detection can be summarized as follows. The traditional phenomenon of efficient temporal integration (integration of signal intensity over about 200 ms) was found to be restricted to a bandwidth of approximately 1/3-oct. With broad-band signals, efficient across-frequency integration was found to be restricted to short signal durations (about 40 ms). This can be illustrated by replotting some of the data obtained with configurations composed of Gaussian-shaped tone pulse components (chapter 2). This pattern (See Figure), showing a 'crosshair', illustrates the two signal conditions which are most favorable for detection: narrow-band long-duration signals for efficient temporal integration, and broad-band short-duration signals for efficient across-frequency integration.

The experiments as presented in the last chapter are concerned with interactions between the signal pattern and the masking noise pattern, with temporally modulated (100% SAM) or spectrally shaped (rippled noise) maskers. An important issue concerns the CMR concept: A release from masking for the detection of a tone pulse in a temporally modulated masker when adding comodulated masking noise remote in frequency from the signal. It was

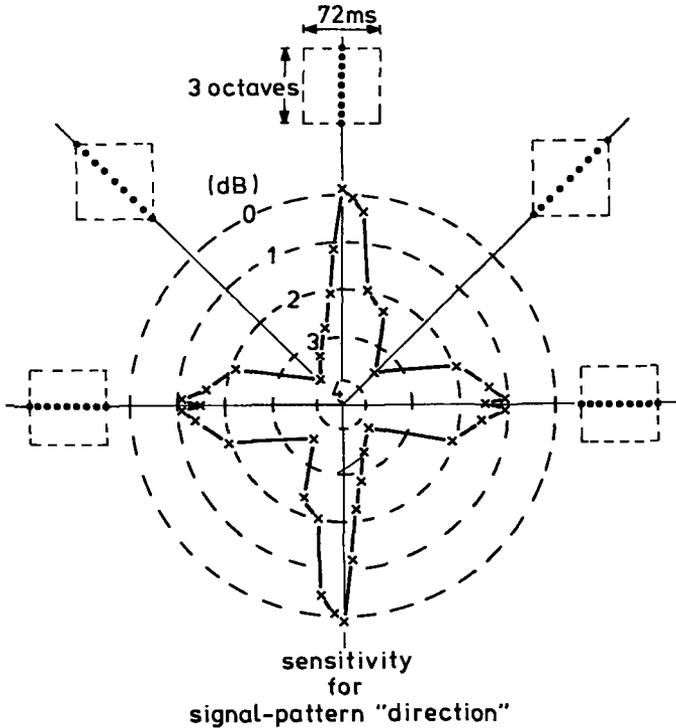


Figure 1. For a given set of signals, each consisting of nine components, as indicated by the schematic diagrams at the border of the figure, the masked thresholds are plotted. The distance to the origin expresses the ear's sensitivity for a certain signal pattern. The ear is most sensitive to click-like and tone-like signals (low masked thresholds).

found, in experiments in which signal bandwidth was varied, that the current models accounting for CMR, based on an across-channel envelope disparity, are inadequate. Effects similar to CMR were found in the time domain; a release from masking for the detection of a narrow-band tone pulse placed in a valley of spectrally shaped noise when adding similarly shaped masker noise preceding and following from the signal. We have called this effect coshaping release from masking (CSMR).

Since a substantial CMR was obtained only for the short-duration signals placed in a masker valley, an interesting consequence is that with increasing noise bandwidth detection threshold for the valley improves, but detection threshold for the peak placement does not improve: The peak-to-valley difference, estimated from detection threshold measurements, depends on the bandwidth of the masker. Similarly, this applies to signal detection in spectrally shaped noise. For the detection of a short-duration narrow band signal, the peak-to-valley difference depends on the duration of the masker.

Thus, the masking effectiveness of a valley, in both temporally modulated noise and spectrally shaped noise, depends on masker bandwidth or masker duration, respectively. This leads to a rather complicated situation when estimating the auditory resolution; it is found that temporal resolution depends on masker bandwidth and frequency resolution depends on masker duration.

It is realized that this thesis is of a phenomenological nature; it provides experimental data on a number of effects related to the spectro-temporal pattern of signal and masker. For the purpose of generalizing these findings to sound perception in general, one would like to derive a functional model of auditory signal detection which does account for the observed phenomena. Numerous attempts have been undertaken in this direction, but at this point the results are not yet considered satisfactory.

ASPECTEN VAN SPECTRO-TEMPORELE

VERWERKING BIJ AUDITIEVE SIGNAALDETECTIE

SAMENVATTING

In deze samenvatting wordt in algemene termen het kader van dit promotie-onderzoek beschreven met een korte beschrijving van een aantal experimenten en resultaten. Er is een aantal intermezzo's ingebouwd om voor de niet-gespecialiseerde lezer een aantal begrippen te verduidelijken.

Het hoofdthema van dit onderzoek betreft de invloed van zowel de signaalpatronen als de stoorlawaaipatronen op de geluiddetectie. Onder een signaalpatroon wordt binnen de context van dit onderzoek de verdeling van de signaalenergie in frequentie en tijd verstaan en onder een stoorlawaaipatroon de verdeling van stoorenergie in frequentie en tijd.

In het alledaagse leven wordt de mens blootgesteld aan een breed scala van geluiden. Ondanks de soms zeer complexe frequentie-tijdpatronen van deze geluiden is de mens verbazingwekkend goed in staat de aangeboden informatie te verwerken. Eén interessant aspect van de auditieve informatieverwerking is het bijzonder goed ontwikkelde vermogen om geluiden te horen tegen een achtergrond van stoorlawaaai. De vraag die hier gesteld wordt is, of, en zo ja, op welke wijze dit vermogen afhangt van het enerzijds het signaalpatroon en anderzijds het achtergrondpatroon? Dit is een vraag die onder laboratorium-condities, gebruik makend van goed gedefinieerde geluiden, bestudeerd kan worden. In deze zogenoemde psycho-akoestische experimenten worden gegevens verkregen die gebaseerd zijn op de reacties van proefpersonen op de aangeboden geluiden. Samen met fysiologische gegevens vormt dat een basis voor een functionele beschrijving van het menselijk auditief systeem.

Intermezzo

De proefpersoon draagt een hoofdtelefoon waardoor continue stoorlawaaai (ruis, ook wel maskeerder genoemd) aangeboden wordt. De detectiedrempel van een signaal wordt bepaald volgens een zogenaamde gedwongen-keuze

methode. Binnen een meetsessie krijgt de proefpersoon herhaaldelijk in één van de twee door lampjes gemarkeerde intervallen naast het stoorlawaaï ook het signaal aangeboden. De proefpersoon moet vervolgens aangeven in welk interval het signaal werd gehoord. Geeft de proefpersoon drie keer achtereenvolgend het correcte interval aan dan wordt het signaal zachter gemaakt, zodra de proefpersoon het verkeerde interval aangeeft wordt het signaal harder gemaakt. Zo wordt het geluidniveau van het signaal bepaald waarbij de proefpersoon in ongeveer acht van de tien keer het juiste interval aangeeft. Dit geluidniveau noemt men dan per definitie de detectiedrempel.

In vrijwel ieder model van het auditief systeem gaat men er vanuit dat men allereerst een frequentieanalyse op het geluid uitvoert. Dit doet men met behulp van een verzameling bandfilters, één voor ieder frequentiekanal. Deze bandfilters blijken een belangrijke rol te spelen bij de auditieve waarneming, niet alleen bij de luidheids-, klankkleur- en toonhoogtewaarneming maar ook bij maskering (voor een overzicht, zie Plomp, 1976 en Moore, 1989). In het licht van de frequentie-selectiviteit van het menselijk oor wordt het geluid vaak gekarakteriseerd door een spectrogram: een patroon waarin het frequentiespectrum van een geluid, bepaald over korte tijdsintervallen, wordt afgebeeld als functie van de tijd. De breedte van de bandfilters die vaak gehanteerd wordt als maat voor de frequentie-selectiviteit van het menselijk gehoor is 1/3-octaf, de kritische band, zoals in veel psycho-akoestische literatuur staat vermeld.

Intermezzo

Een spectrogram is een "plaatje" van geluid, waarbij langs de horizontale as de tijd wordt uitgezet en langs de verticale as de middenfrequenties van de frequentiekanalen. De lokale geluidintensiteit wordt door de hoeveelheid zwarting aangegeven. Een zwart gebied geeft een gebied van hoge geluidintensiteit, een lichtgrijs gebied een gebied van lage intensiteit. Komt het nu voor dat er in verticale richting een heel gebied zwart is dan kan men zeggen dat er in meerdere frequentiekanalen een hoge geluidintensiteit aanwezig is.

In het verleden ging de aandacht met betrekking tot signaaldetectie-experimenten voornamelijk uit naar temporele integratie. Detectiedrempels van zuivere tonen aangeboden tegen een achtergrond van witte ruis werden, als functie van de toonduur, bepaald. In het algemeen werd geconstateerd dat voor een toonduur tot ongeveer 150-200 ms de detectiedrempels bepaald worden door de toonenergie: de toonintensiteit wordt effectief geïntegreerd

over de tijd. Deze experimenten zijn in principe slechts gericht op processen die zich binnen één kritische band afspelen.

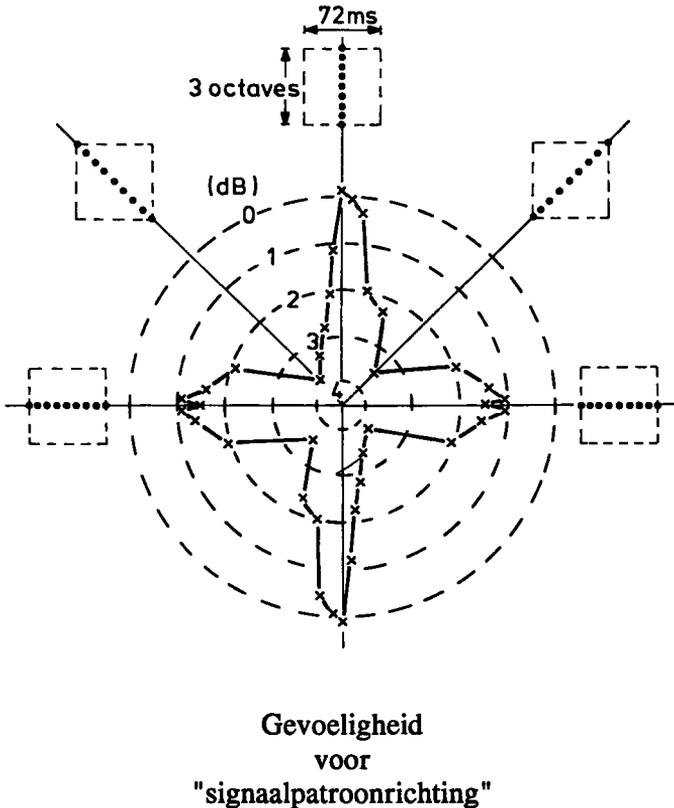
Maakt men echter een spectrogram van natuurlijke geluiden zoals spraak en muziek, dan ziet men dat dergelijke geluiden meerdere kritische banden tegelijk stimuleren. Behalve een integratie van geluidinformatie binnen een kritische band zou dus ook integratie over verschillende kritische banden (spectrale integratie) een rol kunnen spelen bij de geluidwaarneming.

In dit onderzoek wordt een aantal experimenten beschreven waarin systematisch is onderzocht welke rol deze integratie over meerdere frequentiebanden speelt bij de signaaldetectie.

In de literatuur is natuurlijk al het één en ander bekend over deze spectrale integratie. In klassieke experimenten met breedbandige signalen werd gesuggereerd dat het kanaal (frequentie-band) met de beste signaal-ruis verhouding de detectiedrempel bepaalde (e.g. Fletcher, 1940, Gässler, 1954). Dit model werd ook wel het "power-spectrum" model genoemd. Andere experimenten lieten echter zien dat het model niet algemeen geldig was. In experimenten verricht door Green (1958) en Scholl (1962) werd al aangetoond dat er spectrale integratie optreedt bij signaaldetectie. Scholl (1962) gebruikte ruis met een beperkte bandbreedte en vond dat de efficiëntie van de spectrale integratie afhing van de duur van de signalen.

De experimenten die in hoofdstuk 1 en 2 van dit proefschrift beschreven worden zijn gericht op de efficiëntie van de spectrale integratie. Voor een eenduidige interpretatie van spectrale integratie is het vereist dat de detecteerbaarheid van het geluid in ieder afzonderlijk frequentiekanaal gelijk is. In een speciaal hiervoor uitgevoerd experiment werd bepaald hoe het signaal en het achtergrondlawaai (de ruis) op elkaar afgestemd moeten worden om dit te bereiken. Er werd gevonden dat de conditie van gelijke detecteerbaarheid van signaalcomponenten in afzonderlijke frequentiebanden overeen kwam met gelijke $(E/N_0)_{1/3}$ voor elk frequentiekanaal (E geeft de signaalenergie in een 1/3-octaf en N_0 geeft de spectrale dichtheid van de ruis in dezelfde 1/3-octaf) dat gestimuleerd werd door het signaal. Onder bovenstaande randvoorwaarde van gelijke detecteerbaarheid kan men het effect van geluidpatroon op detectie, zoals dat in enkele experimenten werd gevonden, als volgt samenvatten. De traditionele bevinding van efficiënte temporele integratie (integratie van de signaalenergie over ongeveer 200 ms) werd alleen gevonden voor signaalbandbreedten kleiner dan 1/3-octaf. Voor breedbandige signalen werd efficiënte spectrale integratie slechts gevonden voor signalen van korte duur (kleiner dan 40 ms). Dit wordt duidelijk wanneer we enkele experimenteel

verkregen gegevens (uit hoofdstuk 2) opnieuw tekenen. De resulterende figuur hieronder vertoont een "kruisdraad-patroon", en illustreert welke de beste signaalpatronen voor detectie zijn: smalbandige langdurende signalen voor efficiënte temporele integratie en breedbandige kortdurende signalen voor efficiënte spectrale integratie.



Figuur 1. Aan de randen van de figuur zijn schematisch de verschillende frequentie-tijd patronen van de signalen aangegeven. Van deze patronen is in het midden van de figuur de detectiegevoeligheid aangegeven. Zo ziet men dat voor frequentie-tijd patronen waarvoor de signaalcomponenten met gelijke frequentie achter elkaar in de tijd aangeboden worden, en waar de signaalcomponenten met verschillende frequentie tegelijk aangeboden worden, de gevoeligheid voor detectie het grootst is.

Intermezzo

Ten aanzien van het combineren van signaalinformatie kan men dus in het algemeen stellen dat dit bij de detectie van korte signalen even effectief over meerdere frequentiekanalen geschiedt als bij lange signalen binnen één frequentiekanaal over de tijd.

De experimenten die in hoofdstuk 3 zijn beschreven, hebben betrekking op signaaldetectie bij gemoduleerd achtergrondlawaai (het geluidniveau van het achtergrondlawaai fluctueert in de tijd), met als thema de combinatie van twee typen auditieve fenomenen die berusten op mogelijke samenwerking en interactie tussen verschillende frequentiebanden. Naast de hierboven beschreven spectrale integratie bij signaaldetectie wordt in de literatuur een fenomeen beschreven dat ook gebaseerd is op een samenwerking tussen verschillende frequentiebanden: "comodulation masking release (CMR)". In een reeks experimenten verricht door Hall en collega's (1988) werd de detectiedrempel van een zuivere toon in een smal gemoduleerd ruisbandje (gecentreerd rond de signaalfrequentie) bepaald. Vervolgens werd een tweede ruisbandje, met dezelfde modulatie-structuur als het eerste bandje, elders in frequentie aangeboden. Het bleek dat de detectiedrempel van de zuivere toon door toevoeging van het tweede ruisbandje aanzienlijk verlaagd kon worden, mits in beide ruisbandjes het geluidniveau van de ruis op dezelfde wijze fluctueerde. In de modellen die dit verschijnsel proberen te verklaren gaat men er vanuit dat toevoeging van het signaal in een bepaalde frequentieband een verschil in de structuur van de omhullende in deze band t.o.v. die van de andere banden veroorzaakt, en dat op basis hiervan detectie plaats vindt. Een kernpunt is nu de vraag wat er gebeurt wanneer we ook de bandbreedte van het signaal vergroten, zodat meerdere frequentiekanalen tegelijk door het signaal gestimuleerd worden. Volgens de hierboven vermelde modellen zal daardoor een verkleining van het verschil in omhullende-structuur tussen de verschillende frequentiebanden optreden, wat tot een verhoging van de detectiedrempel zou moeten leiden; aan de andere kant zou er misschien een effectieve integratie van signaal-energie optreden (zoals die beschreven in hoofdstuk 1 en 2), zodat de detectiedrempel laag blijft.

Intermezzo

In ons alledaagse leven komen vaak situaties voor waarin we te maken hebben met achtergrondlawaai. Over het algemeen geldt dat het achtergrondlawaai niet mooi constant van niveau is maar dat het aanzienlijk kan fluctueren. Om

een eerste stap te maken in een meer reële maskeersituatie en te onderzoeken hoe detectie zich gedraagt bij zo'n maskeerconditie is gekozen voor een gemoduleerd achtergrondlawaai. In de literatuur over signaaldetectie is al bekend dat bij fluctuerend achtergrondlawaai een fenomeen kan optreden dat ook gebaseerd is op een combinatie van geluidinformatie in de verschillende frequentiekanalen. We willen nu onderzoeken hoe het combineren van informatie in de verschillende frequentiekanalen, zoals onderzocht in hoofdstuk 1 en 2, zich gedraagt samen met een soortgelijk fenomeen, dat in de literatuur bekend staat, als "comodulatie".

Allereerst moet vermeld worden dat een aanzienlijk "comodulatie" effect gevonden werd. Wanneer een kort smalbandig ($1/3$ -octaaf bandbreedte) signaal in een dal van een 100% sinusvormig amplitude-gemoduleerde ruis-maskeerder (SAM-ruis) geplaatst wordt en de ruisbandbreedte wordt vergroot van $1/3$ -octaaf naar $13/3$ -octaaf, dan vindt men een drempelverlaging van ongeveer 14 dB. (hoofdstuk 3, figuur 2). Wanneer nu, bij deze 100% SAM-ruis met een bandbreedte van $13/3$ -octaaf, de signaalbandbreedte wordt vergroot van $1/3$ -octaaf naar $13/3$ -octaaf dan gaat de drempel relatief weinig omhoog (hoofdstuk 3, figuur 4). Dit suggereert dat een vermindering van het verschil tussen de temporele omhullende van de frequentiebanden met signaal en die van de banden zonder signaal niet van doorslaggevende invloed is op het detectieproces.

Tot slot

De hierboven beschreven studie is van fenomenologische aard is. Het zou tot een beter begrip van het functioneren van het menselijk auditief systeem leiden indien de experimentele resultaten onderbouwd zouden kunnen worden met een model. Diverse pogingen in die richting zijn reeds ondernomen maar hebben nog niet tot een bevredigend resultaat geleid.

REFERENCES

Brink, W.A.C. van den, and Houtgast, T. (1988). "Across-frequency integration in short-signal detection", in *Basic Issues in Hearing - Proceedings of the 8th International Symposium on Hearing* edited by H. Duifhuis, J.W. Horst and H.P. Wit (Academic, London).

Brink, W.A.C. van den, and Houtgast, T. (1990). "Efficient across-frequency integration in short-signal detection," *J. Acoust. Soc. Am.* **87**, 284-291.

Brink, W.A.C. van den, and Houtgast, T. (1990). "Spectro-temporal integration in signal detection," *J. Acoust. Soc. Am.* **88**, 1703-1711.

Buus, S. (1985). "Release from masking caused by envelope fluctuations," *J. Acoust. Soc. Am.* **78**, 1958-1965.

Buus, S., Schorer, E., Florentine, M., and Zwicker, E. (1986). "Decision rules in detection of simple and complex tones," *J. Acoust. Soc. Am.* **80**, 1646-1657.

Carlyon, R.P. (1987). "A release from masking by continuous, random, notched noise," *J. Acoust. Soc. Am.* **81**, 418-426.

Carlyon, R.P. (1989). "Changes in the masked threshold of brief tones produced by prior bursts of noise," *Her. Res.* **41**, 223-235.

Cohen, M.F., and Schubert, E.D. (1985). "Place synchrony and the masking-level difference," *J. Acoust. Soc. Am. Suppl.* **1 77**, S49.

Cohen, M.F., and Schubert, E.D. (1987). "Influence of place synchrony on detection of sinusoids," *J. Acoust. Soc. Am.* **81**, 452-458.

Festen, F.M. (1983). *Studies on the relation among auditory functions*, Ph. D. dissertation.

Festen, J.M., Dreschler, W.A.(1988). "Irregularities in the masked threshold of brief tones and filtered clicks," in Basic Issues in Hearing - The 8th International Symposium on Hearing edited by Duifhuis, Horst and Wit (Academic, London).

Fletcher, H. (1940). "Auditory Patterns," *Rev. Mod. Phys.* **12**, 47-61.

Gässler, G.(1954). "Über die Hörschwellen für schallereignisse mit verschiedenem breitem frequenzspektrum," *Acustica* **4**, 408-414.

Green, D.M., Birdsall, T.G., Tanner, W.P.Jr.(1957). "Signal detection as a function of Signal Intensity and Duration," *J. Acoust. Soc. Am.* **29**, 523-531.

Green, D.M.(1958). "Detection of Multiple Component Signals in Noise," *J. Acoust. Soc. Am.* **30**, 904-911.

Green, D.M., McKey, M.J., Licklider, J.C.R.,(1959), " Detection of a Pulsed Sinusoid as a Function of Frequency," *J. Acoust. Soc. Am.* **31**, 1446-1452.

Green, D.M.(1960). "Auditory Detection of a Noise signal," *J. Acoust. Soc. Am.* **32**, 121-131.

Green, D.M. (1973). "Minimum integration time," in *Basic Mechanism in Hearing*, edited by A.R. Møller (Academic Press, New York), 829-843.

Green, D.M., Mason, C.R., and Kidd, G.Jr.(1984). "Profile Analysis: Critical bands and Duration," *J. Acoust. Soc. Am.* **75**, 1163-1167.

Greenwood, D.D. (1961a). "Auditory Masking and the Critical Band," *J. Acoust. Soc. Am.* **33**, 484-502.

Greenwood, D.D. (1961b). "Critical Bandwidth and Frequency Coordinates of the Basilar Membrane," *J. Acoust. Soc. Am.* **33**, 1344-1356.

Grose, J.H., Hall III, J.W.(1988). "Across-frequency processing in temporal gap detection," in Basic Issues in Hearing - The 8th International Symposium on Hearing edited by Duifhuis, Horst and Wit (Academic, London).

- Grose, J.H., and Hall, J.W. (1989). "Comodulation masking release using SAM tonal complex maskers: Effects of modulation depth and signal position," *J. Acoust. Soc. Am.* **85**, 1276-1284.
- Hall, J.W., Haggard, M.P., and Fernandez, M.A.(1984)."Detection in noise by spectro-temporal pattern analysis," *J. Acoust. Soc. Am.* **76**, 50-56.
- Hall, J.W. (1986). "The effect of across-frequency differences in masking level on spectro-temporal pattern analysis," *J. Acoust. Soc. Am.* **79**, 781-787.
- Hall, J.W., and Grose, J.H. (1988). "Comodulation masking release for multi-component signals," *J. Acoust. Soc. Am.* **83**, 677-686.
- Heusden, E. van, and Smoorenburg, G.F.(1981)." Eight-nerve action potentials evoked by tone bursts in cats before and after inducement of acute noise trauma," *Hearing Research* **5**, 1-23.
- Houtgast, T. (1974). *Lateral suppression in hearing*, Ph. D. Dissertation.
- Houtgast,T.(1987)."On the significance of spectral synchrony for signal detection," in *Auditory Processing of Complex Sounds* edited by W.A. Yost and C.S. Watson (LEA, London).
- Levitt, H.(1971)."Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467-477.
- McFadden, D.(1987)."Comodulation detection differences using noise-band signals," *J. Acoust. Soc. Am.* **81**, 1519-1527.
- Moore, B.C.J.(1975)."Mechanisms of masking," *J. Acoust. Soc. Am.* **57**, 391-399.
- Moore, B.C.J., Glasberg, B.R., Plack, C.J., and Biswas, A.K.(1988). "The shape of the ear's temporal window," *J. Acoust. Soc. Am.* **83**, 1102-1116.

Moore, B.C.J., Glasberg, B.R. and Schooneveldt, G.P. (1990). "Across-channel masking and comodulation masking release," *J. Acoust. Soc. Am.* **87**, 1683-1694.

Nabelek, I.V. (1978). "Temporal summation of constant and gliding tones at masked auditory threshold," *J. Acoust. Soc. Am.* **64**, 751-763.

Penner, M.J., and Cudahy, E.(1973). "Critical masking interval: a temporal analog of the critical band," *J. Acoust. Soc. Am.* **54**, 1530-1534.

Penner, M.J., Cudahy, E., and Jenkins, G.W.(1974). "The effect of masker duration on forward and backward masking," *Perception & Psychophysics* **15**, 405-410.

Richards, V.M. (1987). "Monaural envelope correlation perception," *J. Acoust. Soc. Am.* **82**, 1621-1630.

Scholl, H.(1962). "Über die bildung der Hörschwellen und Mithörschwellen von impulsen," *Acustica* **12**, 91-101.

Schooneveldt, G.P., and Moore, B.C.J. (1987). "Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, flanking-band level, and monotic versus dichotic presentation of the flanking band," *J. Acoust. Soc. Am.* **82**, 1944-1956.

Schooneveldt, G.P., and Moore, B.C.J. (1989). "Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth, and signal duration," *J. Acoust. Soc. Am.* **85**, 273-281.

Viemeister, N.F. (1979). " Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.* **66**, 1364-1380.

Viemeister, N.F. (1980). " Adaptation of masking" in *Psychophysical, Physiological and Behavioral Studies in Hearing*, edited by G. van den Brink and F.A. Bilten (Delft U.P., Delft, The Netherlands), 190-197.

Zwicker, E., Feldtkeller, R.(1967). "Frequenzgruppen bei der Verdeckung von Rauschen durch Rauschen" in Das Ohr als Nachrichteneempfänger, Hirzel Stuttgart, 76-78.

Zwislocki, J. (1960). "Theory of temporal auditory summation," J. Acoust. Soc. Am. 32, 1046-1060.

CURRICULUM VITAE

De schrijver van dit proefschrift werd op 20 maart 1958 geboren te Gendt. In 1974 behaalde hij het eindexamen aan de Titus Brandsma MAVO te Arnhem, in 1976 het eindexamen HAVO aan het Nederrijn College te Arnhem en in 1978 het eindexamen Atheneum-B aan het Katholiek Gelders Lyceum te Arnhem. Vervolgens studeerde hij experimentele Natuurkunde aan de Katholieke Universiteit te Nijmegen. Het afstudeerwerk betrof een onderwerp op het gebied van de biofysica; het werd uitgevoerd bij de vakgroep Neurofysica van de afdeling Medische Fysica & Biofysica. Het doctoraalexamen werd op 13 maart 1986 behaald.

Vanaf 15 oktober 1986 tot 15 oktober 1990 was hij in dienst van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) als wetenschappelijk assistent waarbij aan het hier beschreven onderzoek is gewerkt. Het werd uitgevoerd op de afdeling Audiologie van het TNO Instituut voor Zintuigfysiologie te Soesterberg.