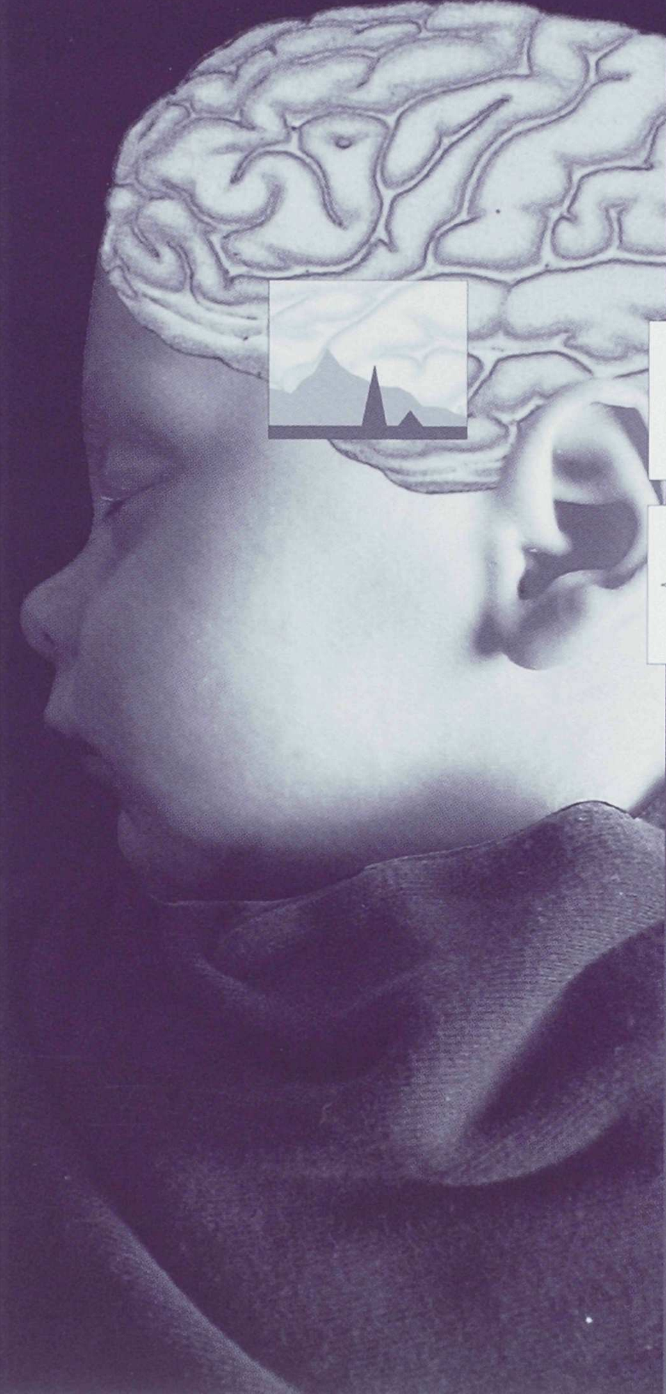
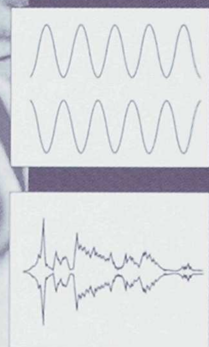
On the perception of sinusoidally amplitude modulated signals and its relevance to listening in noise

Koopman, J.

Publication date
2004

Document Version
Final published version





The perception of sinusoidally amplitude modulated signals and its relevance to listening in noise

Jan Koopman

On the perception of sinusoidally amplitude modulated signals and its relevance to listening in noise

Jan Koopman

Cover design: Andrea Graftdijk.

Printing: Febodruk B.V. - Enschede.

The research described in this thesis has been supported by the Heinsius Houbolt Fund, that funded part of the equipment and the testing facilities.

Printing of this thesis has been supported by Atze Spoor Fonds, Beltone Netherlands B.V., Oticon Nederland B.V., GN ReSound bv, Veenhuis Medical Audio B.V..

On the perception of sinusoidally amplitude modulated signals and its relevance to listening in noise

Academisch Proefschrift

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

Prof. mr. P.F. van der Heijden

ten overstaan van een door het college voor promoties ingestelde

commissie, in het openbaar te verdedigen in de Aula der Universiteit

op dinsdag 16 maart 2004, te 10:00 uur

door Jan Koopman

geboren te Amsterdam

Promotiecommissie:

Promotor : prof. dr. ir. W. A. Dreschler

Co-promotor : prof. dr. ir. T. Houtgast

Overige leden :

prof. dr. E. de Boer

prof. dr. T. Dau

prof. dr. W.J. Fokkens

prof. dr. A. Kohlrausch

prof. dr. ir. H. Spekreijse

prof. dr. G.F. Smoorenburg

Faculteit Geneeskunde

I do not think that the wireless waves I have discovered will have any practical application

Heinrich Rudolf Hertz (1857-94), discoverer of radio wave propagation.

Adapted from :
Scientific blunders, A brief history of how wrong scientists can sometimes be.
by R. Youngson (1998) p.168: published by Robinson - London.

Contents

Chapter 1

Introduction

Dynamic changes in amplitude are important features used in the perception of speech, music and other sounds. The importance of modulations in 'daily-life' signals has led auditory researchers to pay a lot of attention to the perception of sinusoidally amplitude modulated signals (SAM). However, most experiments have measured the modulation detection threshold for SAM. The detection threshold is a measure of sensitivity, and, as such, provides information on the limits of auditory performance. However, the perception of SAM at clearly audible modulation depths is more related to daily auditory tasks. In daily life, the perception of SAM often occurs in the presence of other modulated signals, which is known to reduce the sensitivity to SAM. This reduced sensitivity to SAM is not only known to occur for signals exciting the same auditory filters for target and masker (modulation masking, Houtgast, 1989), but also for conditions in which target and masker carrier excite different filters by choosing clearly different carrier frequencies (modulation detection/discrimination interference or MDI, Yost and Sheft, 1989).

The experimental results described indicate that the perception of SAM occurs at least partly after fusing the outputs of different auditory filters. Modulations are assumed to be an important factor in speech perception. The intelligibility of speech for normal hearing subjects is better when speech is presented in a fluctuating rather than a continuous background noise, when measured at the same signal to noise ratio. This decrease in speech masking effectiveness due to fluctuations in the masking noise is known as *masking release for speech.* When compared to normal hearing subjects, hearing-impaired subjects benefit less or cannot benefit from the fluctuations in the masking noise. This means that masking release for speech is reduced or even absent for hearing-impaired subjects (Festen and Plomp, 1990). Masking release for speech is generally attributed to a number of factors*. Normal hearing subjects may experience less difficulties in segregating the speech from the noise due to a better performance on one of these factors compared to hearing-impaired subjects. However, the perception of speech by hearing-impaired subjects may also be limited by a reduced sensitivity to SAM. Since speech is a highly modulated signal, it may be affected by MDI when presented with modulated maskers. The sensitivity to modulated signals, such as speech, may decrease more when they are presented simultaneously with other modulated signals, such as fluctuating noise, for hearing-impaired subjects than for normal hearing subjects. This thesis focuses on the relation between MDI and masking release for speech. Reduced masking release for speech for hearing-impaired subjects may be caused by excessive modulation masking. To test this hypothesis, the work described in this thesis focuses on three areas:

1. Properties of speech and noise signals: in order to investigate in what ways noises are different from speech, the spectral and temporal properties of noises in a large database were analyzed (Chapter 3).

2. The perception of SAM: MDI was determined for normal hearing and hearing impaired subjects (Chapter 5). In addition, the perception of SAM was determined for highly modulated signals (Chapter 6), since speech is a highly modulated signal and MDI was primarily detected for large modulation depths. These high modulation depths may have obscured the actual relationship between the two parameters, given the difference in modulation perception for low and high modulation depths.

3. Determining the relationships between modulation perception and masking release for speech:

    (a) for simultaneously presented signal and masker (MDI: Chapter 7).

*A number of possible causes of masking release for speech are mentioned in the literature, amongst which co-modulation masking release and glimpsing are seen as most plausible.

(b) for non-simultaneous presentation of signal and masker (adaptation: Chapter 8). Perception may also be affected by stimuli presented before the target (Tansley and Regan, 1979). For modulation perception, this effect may last longer than time constants given in previous studies (*e.g.* forward and backward masking).

(c) at time discrete intervals ranging from simultaneous to non-simultaneous presentation (Chapter 9).

Chapter 2

Introduction to test signals in psychoacoustics

In daily life, the auditory system is exposed to a wealth of acoustical information such as music, speech, and warning sounds. Complex signals like speech or music can be unravelled into a number of signal characteristics, such as level, spectral contents, temporal structure, and phase relations.

The exposure level is an essential parameter in assuring audibility. There are numerous possibilities to express the exposure level. In this thesis, the following methods are used:

1. **dB SPL (Sound Pressure Level);** This is a physical measure of the level and is given by:

   $$dB(SPL) = 10\log_{10}(\frac{I}{I_0}) \tag{2.1}$$

   where I_0 is a reference intensity ($10^{-12}W/m^2$).

2. **dBA (A-weighted);** This is mostly used for broadband signals. The frequencies are weighted according to the weighting of the ear for medium levels.

3. **dB HL (Hearing Level);** This is defined as the level above the average hearing threshold at each frequency for young, healthy listeners with 'normal' hearing.

4. **dB SL (Sensation Level)**; The presentation level of a signal above the subjects' individual threshold of hearing for that particular signal.

2.1 Deterministic signals

The signals discussed throughout this thesis are sinusoidally amplitude modulated (SAM) signals. Figure 2.1 shows two types of sinusoids:

1. the carrier frequency (f_c), given by the fast sinusoid in the left panel of Figure 2.1 with cycle duration [A-B].
2. the modulation frequency (f_m), also referred to as envelope. The envelope is given by the slow sinusoid with offset 1 and cycle duration [A-C] (see left panel Figure 2.1). The dashed sinusoid in Fig. 2.1 is actually a hypothetical one. The modulation frequency connects the maximum values of the time waveform as if there was a sinusoid.

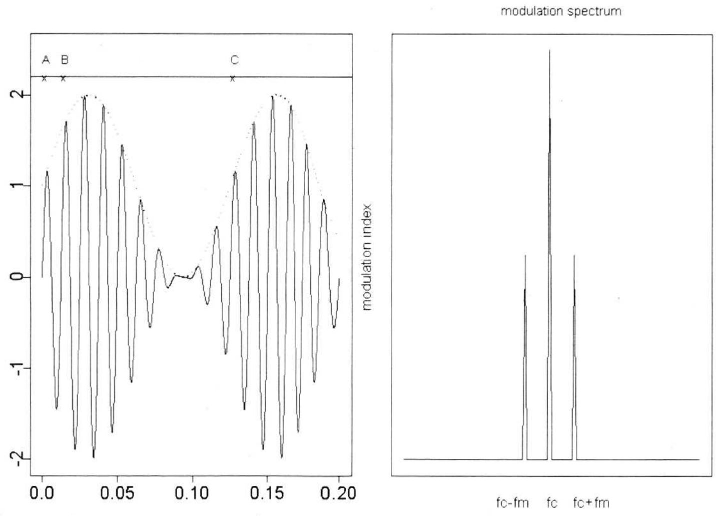


Figure 2.1: *A sinusoidally amplitude modulated signal (SAM) represented in two different ways. The left panel gives the amplitude (ordinate) of a fully modulated (m=1) pure tone as a function of time (abscissa). The right panel gives the spectral representation (abscissa) with the center frequency f_c and the two sidebands (f_c - f_m and f_c + f_m).*

The equation describing this SAM-signal, is given by:

$$\begin{aligned} SAM(t) &= (1 + m\sin(2\pi f_m t + \phi_m))\sin(2\pi f_c t + \phi_c) \\ &= \sin(2\pi f_c t + \phi_c) \\ &\quad + \frac{m}{2}\cos(2\pi(f_c - f_m)t + \phi_c - \phi_m) \\ &\quad\quad + \frac{m}{2}\cos(2\pi(f_c + f_m)t + \phi_c + \phi_m) \end{aligned} \tag{2.2}$$

where the starting phases of carrier and envelope are given by ϕ_c and ϕ_m, respectively. The spectrum of a SAM-signal (right panel Figure 2.1) shows that, apart from the carrier frequency (f_c), two sidebands are directly related to the modulation rate (f_c - f_m and f_c + f_m). The modulation index m determines the strength of the modulation, by a number between 0 (continuous; unmodulated) and 1 (completely fluctuating; modulated), and can be calculated using:

$$m_{lin} = \frac{A_{max} - A_{min}}{A_{max} + A_{min}} \tag{2.3}$$

where A denotes the amplitude of the envelope. For more complex waveform patterns, the amplitude modulation spectrum can be determined by taking the absolute waveform, followed by low-pass filtering in order to eliminate the fine structure, and a FFT (see Figure 2.2*, after which it is divided by the DC-component. For a simple SAM tone, this results in one component in the amplitude modulation frequency domain. This method of extracting the envelope of a signal is very similar to the extraction by the auditory pathway. A more exact, mathematical approach is given by the Hilbert transform†). For high frequencies, both approaches are similar.

*Figure 2.2 is a hypothetical illustration. Carrier and modulation frequencies are generally much higher than illustrated. The cut-off frequency of the low-pass filter in the original STI procedure equalled 35 Hz.

†Since each signal (S_t) can be considered as an analytical, the Hilbert transform (H_t) can be determined according to :

$$H_t = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{S_x}{x - t} dx \tag{2.4}$$

In which the signal and its Hilbert transform have the same power spectrum, the same autocorrelation function, are orthogonal and the Hilbert transform of H_t equals -S_t.
The envelope of the (linear) envelope can now be determined using:

$$L(t) = \sqrt{|S_t|^2 + |H_t|^2} \tag{2.5}$$

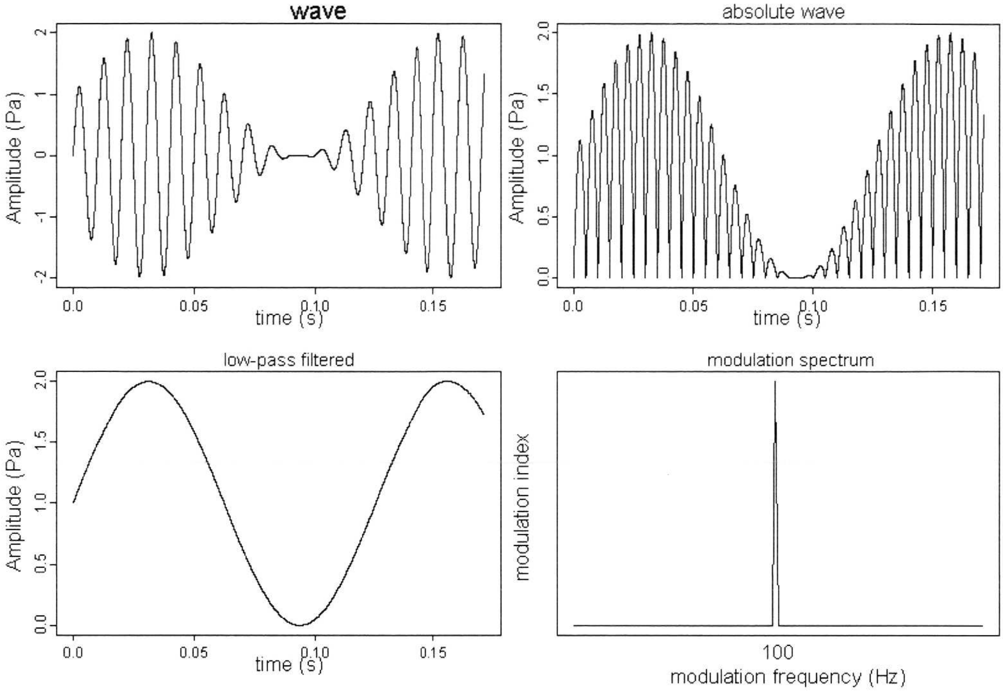


Figure 2.2: *Determination of the modulation spectrum for a SAM-wave. Method is similar to the method described by Houtgast and Steeneken, 1973.*

In psychophysics, the modulation index is often determined by the amplitude waveform (sinusoidally *amplitude* modulated signals). However, the modulation index is also often determined by using the squared waveform. In this case, we usually refer to *intensity* modulation. In that case, the modulation depth is usually determined using:

$$m(dB) = 20 \log_{10} m. \tag{2.6}$$

In order to get a better idea of the temporal fluctuations within a signal, the spectrum of the envelope can be determined.

2.2 Speech

The deterministic signals were developed to emphasize specific aspects of important signals, such as speech. Speech can be described by a number of characteristics which determine the overall waveform. In this section, we will briefly discuss the properties of speech, which are important within the context of this study.

2.2.1 Spectral behavior

When the long term average spectrum of speech (LTASS, Byrne et al., 1994; see Figure 2.3) was studied using speech from 13 different languages and at least 19 speakers for each language[‡], a remarkably high similarity between the speech spectra was observed[§]. The spectra for male (plus-symbols) and female speech (circles) are remarkably similar between 250 Hz and 5000 Hz. Above 6300 Hz female speech contains more energy than male speech, whereas beneath 100 Hz, male speech contains more energy than female speech. Figure 2.3 also indicates that most energy is present below 1 kHz and that above 500 Hz the level of the energy decreases as the frequency increases.

Speech sounds consist of vowels and consonants. The vowels have a clearly pronounced spectrum formed by the formants[¶]. The spectra of the vowels largely determine the LTASS, since the vowels contain more power than the consonants and about half of the time waveform in normal speech consists of vowels. Consonants are usually built up by narrowing or constricting the vocal tract. The transition from consonants to vowels are referred to as formant transitions or co-articulation. In order to visualize these transitions, the spectogram[‖] is used (see Figure 2.4). Dark bands, roughly horizontal, which correspond to the formants, can be seen.

[‡]English(126); Swedish(42); Danish(20); German(27); French(20); Japanese(27); Cantonese(25); Mandarin(21); Russian(21); Welsh(23); Singhalese(21); Vietnamese(19) and Arabic(20)

[§]In a small anechoic room, the speaker was asked to read a piece of text, which was recorded at a distance of 20 cm from the mouth at an azimuth of 45^0 incidence, relative to the axis of the mouth. Sound level meters were set to detect the envelope of speech using a fast time constant (*i.e.* 125 ms).

[¶]the spectral location of the formants are vowel-specific and the lowest three formants are usually below 3 kHz.

[‖]The spectogram is a valuable tool in visualizing the temporal and spectral contents of a signal. However, it is impossible to get a high resolution on the spectral axis and on the temporal axis at the same time. There is a trade-off between them.

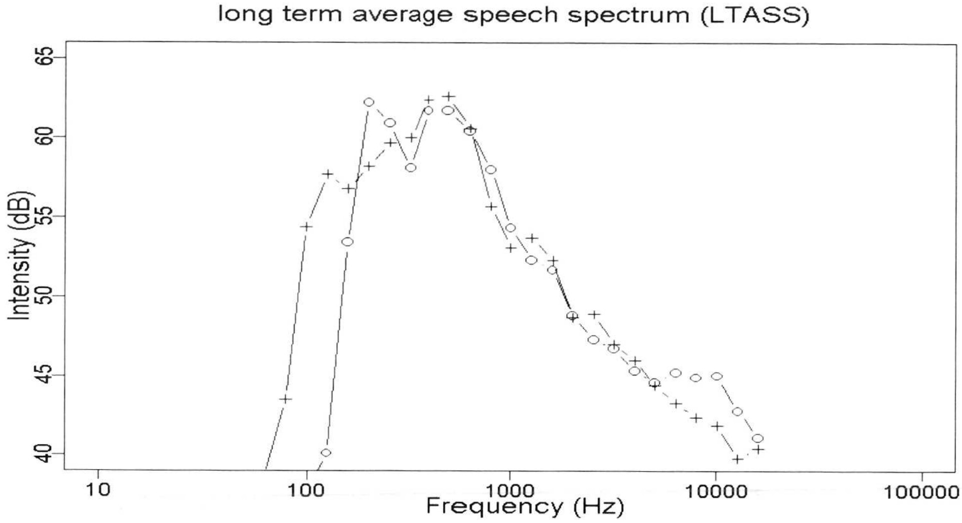


Figure 2.3: *Long term average spectra for male (plus-symbols) and female (circles) speech. There is remarkable similarity between both spectra for the frequency range between 250 Hz and 5000 Hz. Most energy of speech can be found for frequencies between 500 Hz and 1 kHz. Above 6300 Hz female speech contains more energy than male speech, whereas beneath 100 Hz, male speech contains more energy than female speech.*

Redrawn from Byrne et al. (1994)

2.2.2 Temporal behavior

During the last two decades, an increasing amount of research has focused on the temporal properties of sound. The results indicate that the temporal behavior of speech is important to the perception of intonation, rhythm, and the intelligibility of speech. Rosen (1992) formulated a framework for describing the temporal information in sound by distinguishing between three levels

1. **Envelope information******;** The envelope is referred to as the fluctuations in overall amplitude at rates below 50 Hz. This is well within the range given by the envelope of speech (Plomp et al., 1984), since the modulation index for speech is reduced to half its peak value at approximately 15 Hz.

**This definition of the envelope differs from the one given previously based on the analytical signal. In this definition, the envelope contains both envelope and periodicity information.

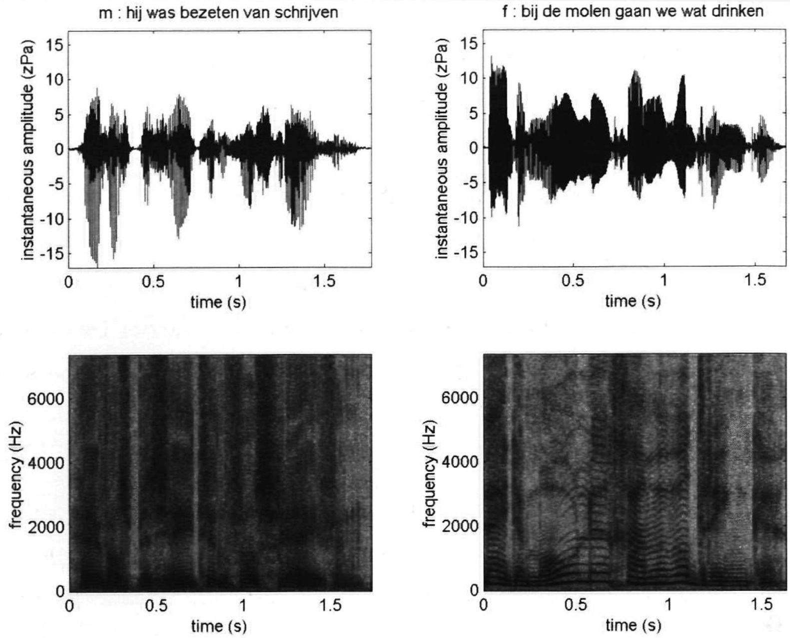


Figure 2.4: *Spectrograms of a sentence uttered by a male (left) and female (right) speaker with the original time waveform in the upper panel and the spectrogram in the lower panel.*

2. **Periodicity information;** Periodicity is referred to as the fluctuation rates between 50 and 500 Hz. Periodic sounds are reflected in the frequency domain by the changes in frequency, leading to a sensation of melodic pitch, which is the main acoustic factor in intonation and accenting syllables.

3. **Fine-structure information;** The fine structure is referred to as the variations between 0.6 and 10 kHz. The fine structure is an important factor in the voice quality, conveys information on the vowel formants, and is important to the intelligibility of very short consonants.

In the scope of this framework, this study focuses on the envelope.

The Envelope of Speech

Speech can be divided into sentences, which can be subdivided into words. The shortest unit of speech is the syllable, which is formed by vowels or consonants. The temporal properties of speech, words and syllables can be identified within the envelope. Figure 2.5 gives the modulation spectrum of speech for octave frequencies between 250 Hz and 4000 Hz determined by the method described in Figure 2.2 (Plomp et al., 1984). Speech consisting of 35 s concatenated sentences was filtered using 4^{th} order Butterworth filters. For each octave band of carrier frequencies, the envelope was obtained by low-pass filtering at 130 Hz and the envelope spectrum was determined. The resulting envelope spectra for the different octave bands show a more or less similar pattern in which the modulations in the high-frequency bands are stronger than in the low-frequency bands. This is probably caused by the fact that vowels contain frequencies within this range. Peaks are usually located around 4 Hz, comparable to the averaged length of a syllable (200 ms). Other typical speech segment durations such as sentences (0.25 Hz), syllables (5 Hz) and phonemes (12 Hz) are also given within the Figure.

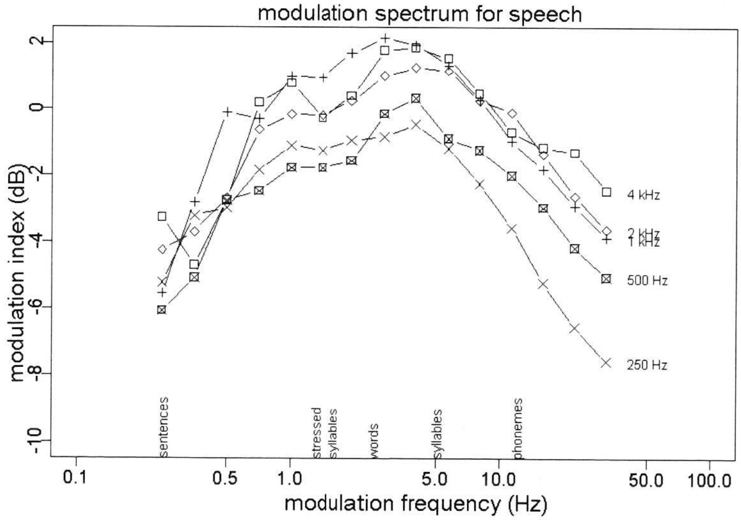


Figure 2.5: *Modulation spectrum of speech for octave bands ranging from 250 Hz to 4 kHz.*

Chapter 3

Towards a representative set of 'real life noises' *

Abstract

The aim of this study is to create a database containing the spectral and temporal behavior of a large number of background noises. The most important dimensions were determined using a factor analysis. This factor analysis reduced the total number of dimensions, by creating a new subspace, from 14 (7 octave bands in the spectral domain and 7 in the temporal domain) to 4. One dimension gives information about the amount of amplitude modulation present in the signal. The other dimensions give spectral information. Based upon the output of the factor analysis, a cluster analysis was carried out to reduce the 144 signals to 15 anchor points. The medians of each cluster form the anchor points. A selection of these anchor points could be used for the fitting of comfort programs in multi-program hearing aids and for the evaluation of signal-processing schemes in specific background noise conditions.

*adapted from Audiology (2001) - 40 pp.78-91

3.1 Motivation

In this study, a database of real-life signals was created. This database may be useful for four reasons.

1. Psychophysical tests, such as those examining speech intelligibility in noise, often use a limited selection of noises. In many situations a simple white noise will not be sufficient to give the researcher a valid representation of problems occurring in real life situations. In speech tests, such as Plomp's adaptive sentence test (Plomp and Mimpen, 1979), a speech-shaped noise is used. A better approximation of 'real life' situations can be achieved using this type of database.

2. Most modern hearing aids have a number of programs implementing several filtering characteristics to reduce the masking effect of specific background noises. Since the specific characteristics of the noises are of great importance, different algorithms can be tuned to these characteristics, which may benefit speech intelligibility (Köbler and Leijon, 1999).

3. It may shed light on acoustical properties important in separating speech from the background noise (see Appendix C).

4. It may provide information regarding the dimensions crucial to speech intelligibility (see Appendix D).

This chapter is divided into two parts. In the first part a large number of signals were collected and their specific properties analyzed. Although level appeared to be a very important factor in the intelligibility of speech (Keidser, 1995), this project focusses on the spectral and temporal behavior of signals. The level is not taken into account, since signals are adapted from CD-ROMs without any information about the level of recording. An important difference with similar work by Keidser (1995) is that the temporal characteristics are taken into account. In the second part of this Chapter, the basic dimensions are determined according to the spectral and temporal behavior of the signals. From this database, dimensions can be determined that explain most of the variance between the signals. These dimensions can be regarded as the axes of a multi-dimensional subspace representing the most important characteristics of the set of background noises. Following the determination of these basic dimensions, a number of 'anchor signals' were selected. Within the multi-dimensional space, these 'anchor signals' are distributed as widely as possible and, hence, form an accurate representation of the whole space given by all noises.

3.2 Methods

3.2.1 Selecting the signals for the database

There were some restrictions on the selection of signals. Stable or fluctuating environmental signals were included, while intermittent and impulsive signals were excluded since they would lead to uninformative modulation spectra.

Sound category	**Number of signals**	**Identification numbers of the signals**	**Symbol**
Noises at work	12	1-12	$+$
Noises in and around the car	12	13-24	$*$
Speech in noise	10	25-34	$\diamond$
Speech signals	11	35-45	$\times$
Noises at home	12	46-57	$\bullet$
Environmental noises	10	58-67	$\circ$
Army noises	10	68-77	∇
City noises	10	78-87	$\square$
Crowds indoor	12	88-99	$\triangle$
Hobby sounds	11	100-110	$\triangleleft$
Music	12	111-122	$\triangleright$
Transport sounds	10	123-132	%
Nature sounds	12	133-144	#

Table 3.1: *Sound categories with appropriate signal numbers and symbols*

In total 144 signals were selected from commercially available CDs[†] and the authors' own recordings[‡]. The signals were divided into 13 categories, given in Table 3.1, to avoid over representation of certain types of signals. A short description and the original CD title and track number of each signal is given in Table 3.2.

Signal database

	Description	Origin	Track			Description	Origin	Track	
N1	Industry	Widex	8	+	N73	Leopard1	TNO	13	∇
N2	Office	S.I.	14	+	N74	HF radio	TNO	17	∇
N3	Bandsaw	Widex	12	+	N75	F16	TNO	20	∇
N4	Compressed air	Widex	13	+	N76	Destroyer (engine)	TNO	15	∇
N5	Industry (cont.)	Widex	9	+	N77	Buccaneer	TNO	8	∇
N6	Printing industry	Widex	14	+	N78	City	S.I.	1-15	□
N7	Factory for cars	TNO	22	+	N79	Crowd3	S.I.	7-12	□
N8	Plate cut factory	TNO	21	+	N80	Crowd2	S.I.	7-7	□
N9	Office light	S.I.	10-24	+	N81	Crowd1	S.I.	7-1	□
N10	Fan	S.I.	12-4	+	N82	City light	S.I.	1-3	□
N11	Construction city	S.I.	2-9	+	N83	Harbour city	S.I.	2-15	□
N12	Building site	Siemens	53	+	N84	Park city	S.I.	2-5	□
N13	Highway	S.I.	5-7	∗	N85	Crowd outdoor	S.I.	7-10	□
N14	Traffic in car	Siemens	52	∗	N86	Crowd outdoor	S.I.	7-14	□
N15	Traffic in street	WIDEX	6	∗	N87	Traffic	S.I.	11-1	□
N16	Volvo bricks	TNO	23	∗	N88	Crowd indoor	S.I.	9-11	△
N17	Volvo asphalt	TNO	24	∗	N89	College-lounge	S.I.	9-7	△
N18	Inside car	WIDEX	7	∗	N90	Bar	S.I.	10-12	△
N19	Inside car	Siemens	8	∗	N91	Restaurant	S.I.	10-6	△
N20	On the road traffic	Multifocus	7	∗	N92	Indoor party	WIDEX	4	△
N21	Car 100km/h	o.r.	n.a.	∗	N93	Bureau office	WIDEX	10	△
N22	Car 80km/h	o.r.	n.a.	∗	N94	Talking Office	WIDEX	11	△
N23	Quiet highway	o.r.	n.a.	∗	N95	Canteen	TNO	19	△
N24	Petrol station	multifocus	6	∗	N96	Restaurant	S.I.	10-1	△
N25	1♂, 1♀ (artificial)	ICRA	6	◇	N97	Bar	S.I.	10-10	△
N26	♀,♂;2(♀,♂)@-6 dB	ICRA	7	◇	N98	Crowd indoor	S.I.	6-1	△
N27	N26 but loud	ICRA	9	◇	N99	Crowd indoor	S.I.	8-10	△
N28	Babble SNR_{low}	Phonak	7	◇	N100	Football stadium	Multifocus	11	◁
N29	Cafeteria SNR_{low}	Phonak	5	◇	N101	AJAX stadium*	o.r.	n.a.	◁
N30	" " SNR_{high}	Phonak	4	◇	N102	Excited stadium**	o.r.	n.a.	◁
N31	Walking & talking	o.r.	n.a.	◇	N103	Quiet stadium**	o.r.	n.a.	◁
N32	Babble - low SNR	Phonak	7	◇	N104	Horse riding	o.r.	n.a.	◁
N33	♀ restaurant	H.J.	15	◇	N105	Entering stadium	o.r.	n.a.	◁
N34	♀ museum	H.J.	13	◇	N106	Applause	o.r.	n.a.	◁
N35	♀ talker	ICRA	4	×	N107	Live Classical music	o.r.	n.a.	◁

continued on next page

[†]Signals were adapted from the following CDs; Plomp and Mimpen 1979; Festen and Plomp 1990; Dreschler et al. 2001, Widex, 1995®,Oticon, 1993®, Siemens, 1992®, Hearing Journal, 1998, TNO, 1990®, Phonak, 1993®, Sound Ideas, 1988®, Oticon, 1993®, Brüel & Kjær, 1989®.

[‡]The recordings were made using a Stereo condenser electret microphone Philips SBC3050.

continued from previous page

	Description	Origin	Track			Description	Origin	Track	
N36	♂ talker	ICRA	5	×	N108	Swimming pool	o.r.	n.a.	◁
N37	Speech-noise	TNO	6	×	N109	Rehearsing concert hall	o.r.	n.a.	◁
N38	♂ sp.sh. noise$_{fluc}$	Plomp		×	N110	TV	o.r.	n.a.	◁
N39	♂ sp.sh. noise$_{cont}$	Plomp		×	N111	Opera	Zauberflöte	15	▷
N40	♀ sp.sh. noise$_{fluc}$	Plomp		×	N112	Choir	Desprez	6	▷
N41	♀ sp.sh. noise$_{cont}$	Plomp		×	N113	Classical music	Brahms	4	▷
N42	CCITT	phonak	2	×	N114	House kids	house	8	▷
N43	Speech babble	phonak	40	×	N115	South american	Simon	1	▷
N44	Cafeteria	phonak	39	×	N116	Drum	Kampen	2	▷
N45	Hans Boer	H.B.	n.a.	×	N117	Country	deLange	1	▷
N46	Toilet	o.r.	n.a.	•	N118	Blues	Dan	1	▷
N47	Hood	o.r.	n.a.	•	N119	Jazz	Holiday	1	▷
N48	Buzzing fridge	o.r.	n.a.	•	N120	Port. folk music	Rodriquez	1	▷
N49	Geyser	o.r.	n.a.	•	N121	African music	Kidjo	1	▷
N50	Microwave	o.r.	n.a.	•	N122	Reggae	Cliff	1	▷
N51	Telephone	o.r.	n.a.	•	N123	Wheel carrier	TNO	10	%
N52	Drier	o.r.	n.a.	•	N124	In steamtrain	Multifocus	2	%
N53	Hair drier	o.r.	n.a.	•	N125	Slow train	o.r.	n.a.	%
N54	Vacuum cleaner	o.r.	n.a.	•	N126	Under ground	o.r.	n.a.	%
N55	Steady vacuum	o.r.	n.a.	•	N127	Train	o.r.	n.a.	%
N56	Washing machine	o.r.	n.a.	•	N128	Train in tunnel	o.r.	n.a.	%
N57	Shower	o.r.	n.a.	•	N129	Fast train	o.r.	n.a.	%
N58	Country	S.I.	4-4	○	N130	Locomotive	Siemens	51	%
N59	Kids	S.I.	7-17	○	N131	Trains arriving	Multifocus	5	%
N60	Market	S.I.	9-17	○	N132	New tram+speech	o.r.	n.a.	%
N61	Large party	Widex	4	○	N133	Crickets	S.I.	4-6	#
N62	Small party	Widex	5	○	N134	Thunderstorm	Multifocus	10	#
N63	Passing train	Brüel & Kjær	18	○	N135	Jungle day	S.I.	3-2	#
N64	Airport exterior	S.I.	5-15	○	N136	Ocean and birds	new age	1	#
N65	Bus-station	S.I.	9-10	○	N137	River frog cricket	S.I.	4-1	#
N66	Train-station	S.I.	9-4	○	N138	Desert night	S.I.	3-9	#
N67	Fun fair	Siemens	55	○	N139	Swamp night	S.I.	3-7	#
N68	Tank cockpit	TNO	14	▽	N140	Desert day	S.I.	3-8	#
N69	Machine gun	TNO	16	▽	N141	Forest day	S.I.	3-13	#
N70	M109	TNO	7	▽	N142	Mountain day	S.I.	3-11	#
N71	LYNX	TNO	12	▽	N143	Birds in nature	S.I.	3-3	#
N72	Leopard2	TNO	9	▽	N144	At the beach plane	Multifocus	4	#

Table 3.2: *The signals included in the database. All signals were recorded from the first second on. The abbreviations o.r. and S.I. stand for 'own recording' and 'Sound Ideas', respectively.*

* *indoor football stadium*

** *indoor speed skating stadium Thialf*

3.2.2 Acoustical analysis

The temporal and spectral behavior of the signals was analyzed using a very similar procedure as the one described by Verschuure et al. (1998). The same segments were used for both analyzes.

Frequency analysis

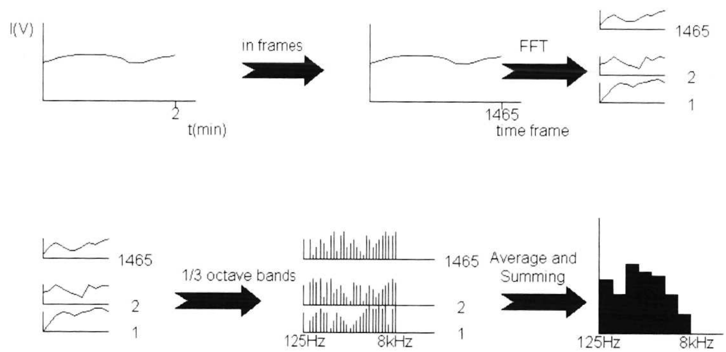


Figure 3.1: *Spectral analysis of a signal lasting two minutes. The signal is divided into 1465 different temporal frames. Partitioning these frames into octave bands by averaging the power within one third octave bands resulted in the octave band spectrum.*

Figure 3.1 illustrates the frequency analysis was carried out for a two-minute segment. Each segment was sampled at a 25 kHz sample rate, resulting in 1465 50%-overlapping frames of 4096 samples. The 1465 (163.8 ms) time signals were Hanning-windowed and Fourier transformed (FFT) to determine the frequency spectrum. After the determination of the FFT, the spectrum was divided into 17 one-third octave bands with central frequencies ranging from 125 Hz to 8 kHz. This one-third octave band spectrum was transformed into an octave band spectrum using intensity summation in order to retrieve an equal number of parameters for the spectral and temporal domain. For the lowest and the highest octaves the levels were estimated from the two other one-third octave bands available within that band.

Temporal analysis

Figure 3.2 shows how the envelope of the unfiltered signal is determined by squaring the signal (intensity), low-pass filtering ($f_c = 100$ Hz) and downsampling ($f_c = 1200$ Hz) (Houtgast and Steeneken, 1973). When the envelope had been obtained, the spectrum was calculated using a FFT and values were summed within 8 octaves (reaching from 0.5 Hz to 64 Hz as indicated by Houtgast, 1989). The modulation index was calculated by dividing the energy in the octave filter by the DC component (given by the first component of the FFT; 0 Hz).

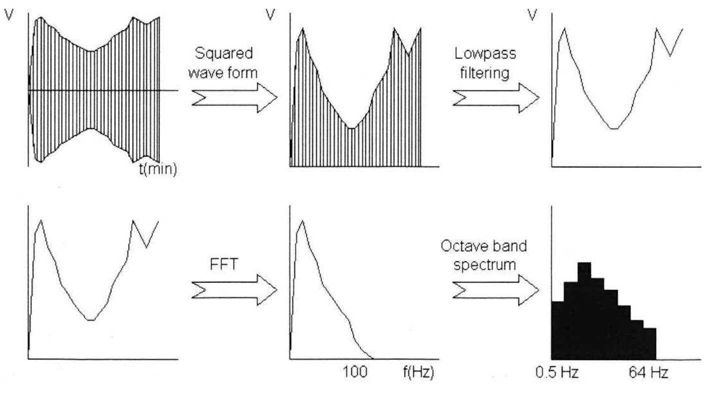


Figure 3.2: *The spectrum of the envelope was obtained by taking the squared waveform of the signal and low-pass filtering. From this envelope, the FFT was calculated and subdivided in different octave bands (ranging from 0.5 to 64 Hz). Dividing these averages by the estimated DC-component resulted in the modulation spectrum.*

3.2.3 Data reduction

The parameters resulting from the analysis of the signals were examined using SPSS 10.0. A factor analysis (principal axis factoring) was carried out and the factors explaining the variance of the spectral and temporal bands were determined. Since most signals were adapted from a number of CDs, there was no information on the level of recording. Therefore, all spectra were scaled to 70 dBA and the input for the factor analysis is given by the spectra in dB SPL. This normalization prevented the overall intensity becoming a factor in the PCA. Although hearing-impaired subjects will probably react differently to signals as a function of the exposure level, it was not possible to include the level of the signals. If level information were available, this information could easily be combined with the spectro-temporal properties investigated in this study. The temporal analysis resulted in 7 octave bands and the spectral analysis in 17 one-third-octave bands. The latter was reduced to 7 octave bands in order to balance the influence of both aspects. The factor analysis was carried out on the 14 $\times$ 14 correlation matrix between the 7 octave bands (dB SPL) and 7 modulation indices.

3.2.4 Choice of anchor points

As was mentioned in the Motivation of this study, the reason for using multidimensional scaling was to find representative anchor points. The anchor points were determined using a cluster analysis (hierarchical clustering, median clustering, squared Euclidean distance using SPSS 10.0). The cluster analysis reduced the number of signals. For the final number of anchor points, a compromise had to be found between a good representation of all signals within the field and a practical number of anchor points to obtain a reasonable representation. For each cluster, the squared Euclidean distance between pairs of clusters in the space defined by the dimensions obtained from the factor analysis was determined. The location of a cluster was determined by the median of the signals within a particular cluster. The Eigenvalue, a representation of the importance of a factor belonging to this dimension, was the multiplying factor for a weighted distance on each axis. The first stage in the cluster analysis was calculating the distances between all signals (each forming their own cluster), the two signals (clusters) that are closest are joined into a new cluster (signal 1 and 2). Then a new cluster (signal 3) joined the cluster already formed if the minimum distance of signal 3 to the median of the cluster (formed by signal 1 and 2) was smaller than the distance between signal 3 and another cluster (signal 4). Otherwise, the two closest clusters (signal 3 and 4) were placed in a new cluster. This procedure was repeated until all signals are placed in an a-priori specified number of clusters.

3.3 Results

3.3.1 The factor analysis

The number of factors, which contribute to the explained variance, can be determined in two ways:

1. the number of Eigenvalues that exceed the average Eigenvalue.

2. the number of Eigenvalues before a knee point in the plot of Eigenvalues (*i.e.* Scree plot).

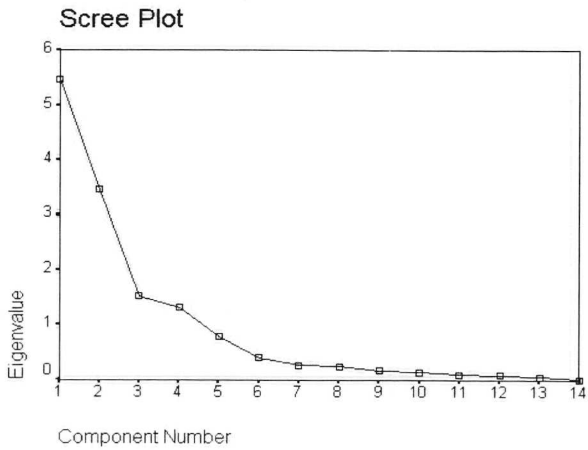


Figure 3.3: *Scree plot with the averaged Eigenvalue as a function of the variable number.*

Figure 3.3 presents a Scree plot with the normalized Eigenvalue on the ordinate as a function of the factor from the factor analysis. According to this Figure, the first four factors contribute to the overall explained variance. Following a Varimax rotation, the first four factors explain 84.1% of the variance (Kleinbaum et al., 1987).

factor nr.	variance explained (%)
1	39.1%
2	24.7%
3	10.9%
4	9.4%

Table 3.3: *Variance explained by each factor*

All four factors can be described by a specific filtering characteristic in either the temporal or the spectral domain (Figure 3.4). The first factor for instance, describes the total amount of amplitude modulation present (explaining 39% of the total variance). The second factor explains an additional 25% of the total variance using a low-pass filtering characteristic followed by negative weighting around 2 kHz (called a band-stop characteristic). The third factor represents a bandpass filter around 4kHz (explaining 11% of the total variance) and the fourth represents a bandpass signal with 1 kHz as its center frequency (explaining 10% of the total variance).

All signals have their specific place in the four dimensional space defined by the factor analysis. In Figure 3.5 the factor loadings have been presented in two-dimensional figures with a separate symbol for each category (see Table 3.1)

3.3.2 Anchor points

As mentioned, the anchor points have been determined using a cluster analysis on the coordinates of the signals multiplied by their Eigenvalue. Each cluster is represented by the signal forming the median of the cluster. The 15 anchor points are given in Table 3.4.

It can be seen that the collection of anchor points covers a wide range of background signals to which the hearing aids and the communication devices can

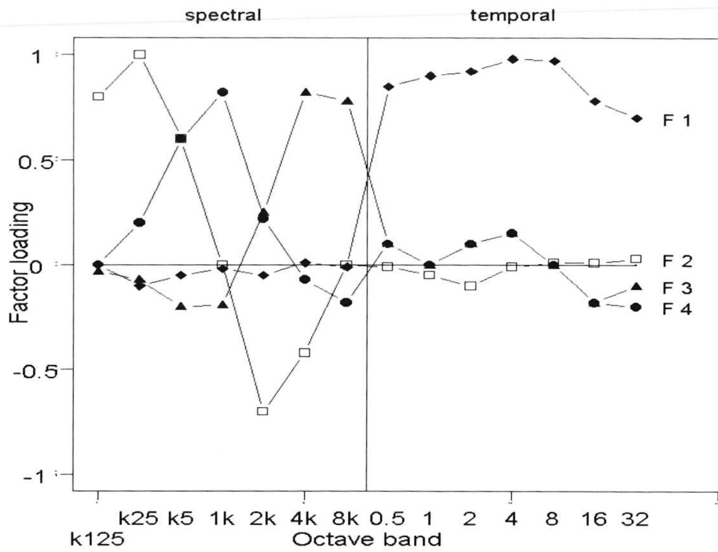


Figure 3.4: *The 4 significant factors given in the original 14 dimensional space.*

be fitted to. For instance, the group of speech signals is represented by the ICRA noise. Other categories, such as crowds indoors or noises at work, also cover common background signals. A categorization of the spectral and temporal behavior for each cluster is given in Table 3.5. The spectral analysis is subdivided into three different types of filtering: low-pass, bandpass and high-pass. The temporal activity, defined as the amount of amplitude modulation present (m), is split into: almost absent (0: $m < 0.2$); weak (*: $0.2 < m < 1.2$); moderately strong (**: $1.2 < m < 1.8$); and very strong (***: $m > 1.8$). The modulation indices are averaged over bands and the range was determined.

Scatter plots of the signals are given in Figure 3.5. The axes are the dimensions defined by the factor analysis and displayed in Figure 3.4. The signals given by the symbols (see Table 3.1), and clusters given by the letters are well spread out along the four dimensions.

Cluster letter	Signal	Situation	Category	Track
A	N1	Industry	noises at work	Widex6- 8
B	N33	Restaurant + ♀ talker	speech in noise	Hearing Journal-15
C	N3	Bandsaw	noises at work	Widex6- 12
D	N106	Applause	hobby sounds	n.a.
E	N12	Building site	noises at work	Siemens-53
F	N14	Traffic in car	car noises	Siemens-52
G	N23	Quiet highway	car noises	n.a.
H	N100	Football stadium	hobby sounds	Multifocus-11
I	N35	♀ talker	speech signals	ICRA-4
J	N133	Crickets	nature sounds	Sound Ideas 4-6
K	N69	Machine gun	army noises	TNO-16
L	N94	Talking office	crowds indoor	Widex-11
M	N134	Thunderstorm	nature sounds	Multifocus-10
N	N135	Jungle day	nature sounds	Sound Ideas 3-2
O	N139	Swamp night	nature sounds	Sound Ideas 3-7

Table 3.4: *Fifteen clusters with their median signal. The selection was based on a cluster analysis. In the first column, the letter is given which indicates the cluster (in Figure 3.5). The number of the signal (column 2) and the description of the signal (column 3) is given for the median of the cluster. The final column gives the origin of the track.*

3.4 Discussion

In this project, 144 signals have been analyzed to create a multi-dimensional space to describe over 80% of the variance in the spectral and temporal behavior. The temporal behavior was determined by the modulation spectrum (Houtgast and Steeneken, 1973). The spectral behavior was determined in seven octave bands and were calculated by adding the energies across three one-third octave bands. The factor analysis gives a good impression of the factors, which can be used to explain most of the variance among signals. The spectral content is well represented as three factors describe the spectral behavior of all signals. However, only one factor describes the variance of the temporal behavior. This suggests that information on the shape of the modulation spectrum is not the primary

Number of signals	Cluster	Spectral low-pass	bandpass slope-LF	slope-HF	high-pass	temporal activity (m)
13	A		flat	flat		0
79	B	-3 dB/oct				0 - *
2	C		2kHz: -7dB/oct	2kHz: 9dB/oct		*
3	D				5 dB/oct	*
12	E		2kHz: flat	2kHz: 6dB/oct		* - **
6	F	-6dB/oct				*
1	G		2kHz: flat	2kHz: 10dB/oct		***
16	H		1kHz: -3dB/oct	1kHz: 9dB/oct		*
4	I		flat	flat		* - **
3	J		flat	flat		*
1	K	-9dB/oct				**
1	L		flat	flat		** - ***
1	M	-3 dB/oct				***
1	N		2kHz: -13dB/oct	2kHz: 9dB/oct		*
1	O		2kHz: -12dB/oct	2kHz: 15dB/oct		* - **

Table 3.5: *The characteristics of all clusters based upon their spectral and temporal behavior. Behavior is specified using spectral and temporal properties. The spectral behavior was divided into low-pass, bandpass (LF; low frequency side, HF; high frequency side) and high-pass characteristics with additional spectral slopes. The temporal behavior was classified according to the strength of the modulation;* 0; $m < 0.2$; * $0.2 < m < 1.0$; ** $1.0 < m < 1.8$; *** $m > 1.8$.

characteristic to separate the signals in this database. However, more details about the temporal characteristics may be useful in some research settings. In this case, a separate principal components analysis could be applied on the temporal domain (see Appendix A).

Keidser (1995) showed that 'city dwelling people' are mostly exposed to low frequency background signals, and that conversation at 'casual' levels in these environments will cause great difficulties. However, not all daily-life signals can be referred to as a low-frequency background noise.

The use of the database is two-fold. On the one hand signals can be selected for very specific purposes, for instance audiological testing. On the other hand the database can be used as a reference source. In order to use the database as a reference source, measurements would be needed on speech intelligibility or different

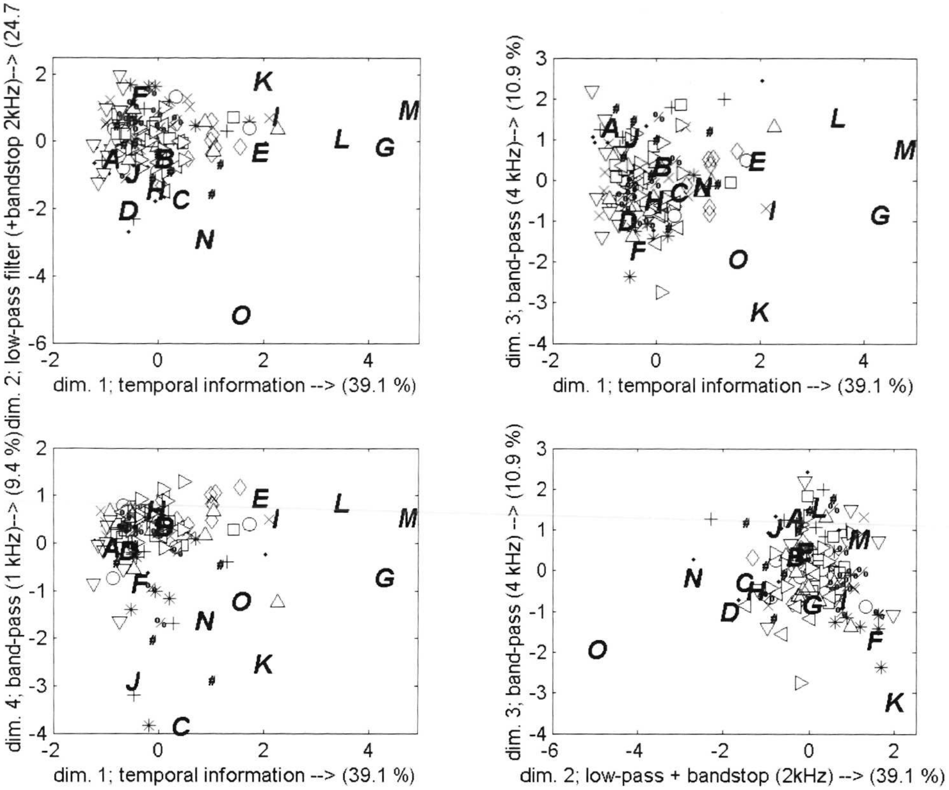


Figure 3.5: *Factor loadings for the four-dimensional space for the 144 different signals. Each signal is denoted by a symbol (see Table 3.1). The signals representing clusters are denoted by letters (see Table 3.5).*

algorithm settings for various noise-reduction strategies. These data can be useful to tune different settings to the specific demands of the subject. Ideally the subject would receive a tape-recorder from the audiological center, to make recordings of frequently occurring situations. These recordings can be analyzed in a similar way. After the determination of the four factor loadings for these recordings, the closest cluster to this signal contains the reference data. The settings for the hearing aid or communication equipment would then be adapted according to the reference data belonging to this cluster (for calculation routines see Appendix B).

3.5 Conclusions

The factorial analysis showed that four factors were sufficient to explain 84% of the variance in the data. The most important factor represents the amount of amplitude modulation present in a signal. The other three factors represent the spectral properties of a sound. These are a low-pass coupled to a band-stop filter (f_c=2 kHz), a bandpass filter (f_c=4 kHz) and a bandpass filter (f_c=1 kHz). Anchor points have been determined using a cluster analysis. In total, 15 clusters of signals have been determined, representing the database as a whole with their temporal and spectral properties. These anchor points can be used for fitting of comfort programs in multi-program hearing aids and for the evaluation of signal-processing schemes in specific background noise conditions.

A closer examination of the signals in Appendix C, illustrates that speech and noise signals deviate predominately in their temporal characteristics. This temporal characteristic is very well used by normal hearing subjects in order to improve speech intelligibility, but cannot be used by hearing-impaired subjects (see Appendix D).

3.6 Appendix A - Temporal properties

In order to select signals on the basis of their temporal behavior, it may be relevant to know more about the details of the amplitude modulation spectrum. Therefore, the most important factors in the amplitude modulation spectrum were determined using a factor analysis. The Scree plot results in two significant factors explaining most of the variance. The factor loadings for these two factors are given in Figure 3.6.

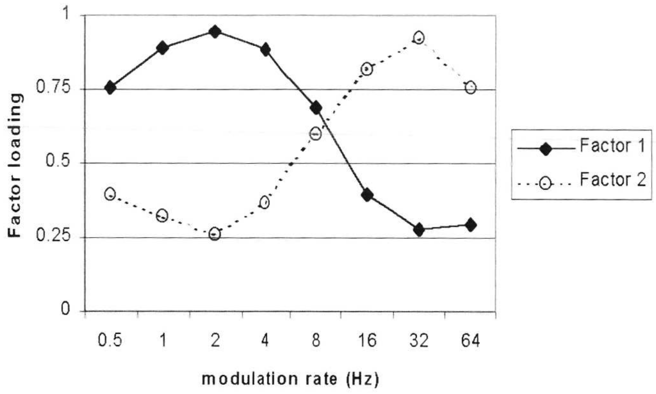


Figure 3.6: *The two dimensions explaining most of the variance after Varimax rotation. The first dimension predominantly shows the presence of slow amplitude modulations and the second dimension the presence of fast amplitude modulations*

Figure 3.6 shows that the first dimension, explaining 74% of the variance, loads principally on the lower modulation rates. The second dimension loads principally on the higher modulation rates. Plotting these factor loadings in a two-dimensional plane results in Figure 3.7. The cluster analysis identified 10 clusters. The signals given in Table 3.6 are indicated by letters in Figure 3.7 and give a clear idea of the importance of the low and high amplitude modulation frequencies.

Signal	Cluster	Name	Average MI	MI f <= 4Hz	MI f > $4Hz$
N45	A	Hans Boer	0.33	0.21	0.45
N132	B	New tram & speech babble	0.53	0.50	0.56
N12	C	Building site	1.01	1.29	0.79
N23	D	Quiet highway	2.12	1.76	2.47
N28	E	Babble (Engl.) -low SNR	0.94	1.09	0.80
N97	F	Bar	0.85	0.89	0.82
N93	G	Bureau office	1.29	1.66	0.92
N69	H	Machine gun	1.41	0.94	1.89
N94	I	Talking office	1.93	2.2	1.66
N133	J	Crickets	0.46	0.07	0.87

Table 3.6: *Description of the 10 clusters in terms of temporal behavior.*

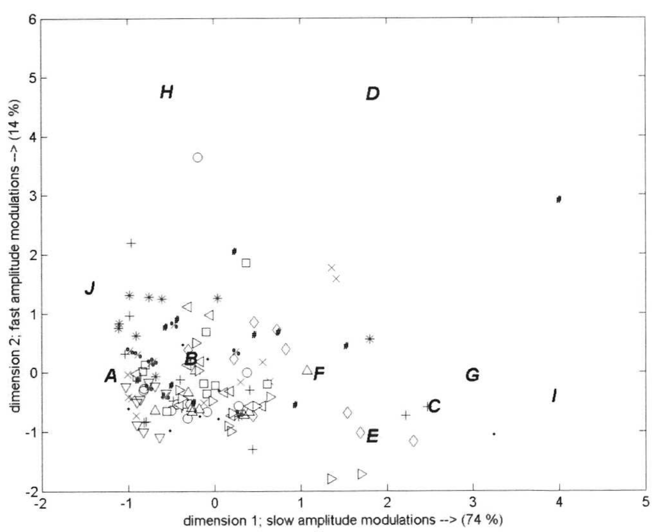


Figure 3.7: *Factor loading plot for all signals in the modulation domain. Symbols represent each category (see Table 3.2) and clusters are identified by the letters (Table 3.6). The first dimension represents lower modulation rates and the second dimension represents the signals with higher modulation rates.*

3.7 Appendix B - Placing signals in the database

The Motivation of this study described an ideal situation, where the subject would identify situations in which he or she is most bothered by background noise. This signal would be recorded by the subject and introduced into the fitting procedure.

In general, it would be difficult to use these signals in the fitting procedure or to

	average	standard deviation
125 Hz(f)	60.96	11.66
250 Hz(f)	64.15	8.04
500 Hz(f)	65.19	4.83
1 kHz(f)	63.50	3.42
2 kHz(f)	61.54	3.78
4 kHz(f)	56.55	5.19
8 kHz(f)	48.29	8.26
0.5 Hz(t)	0.448	0.382
1 Hz(t)	0.502	0.436
2 Hz(t)	0.528	0.432
4 Hz(t)	0.577	0.467
8 Hz(t)	0.555	0.396
16 Hz(t)	0.572	0.407
32 Hz(t)	0.615	0.422
64 Hz(t)	0.677	0.395

Table 3.7: *Average and standard deviation for all 15 factors: spectral (f) and temporal (t)*

adapt standard settings of hearing aids for these signals. This database of signals has been developed in order to meet this problem. A large number of signals has been analyzed to examine the structure of the signals and anchor points are determined in order to symbolize the most representative signals. To use this database in the future, it is necessary to be able to determine the location of a new signal within the new 4-dimensional subspace (factor-analysis) or the 2-dimensional subspace (temporal).

Firstly, the spectra have to be normalized to 70 dBA and transformed into dB SPL. The input of the factor analyzes is a normalized database. The averages and standard deviations for the different parameters are given in Table 3.7. In order to determine the factor loadings of a specific signal, the data have to be normalized by subtracting the average (over the 144 signals) for each parameter from the parameter itself and dividing the outcome by the standard deviation (over the 144 signals). This results in a matrix B (Table 3.7) for the combined analysis. a [14×1] matrix.

original	Spectro-temporal Analysis				Temporal Analysis	
dimension	factor 1	factor 2	factor 3	factor 4	factor 1	factor 2
125 Hz (f)	0.024	-0.015	-0.010	-0.070	-	-
250 Hz (f)	-0.003	1.020	0.413	0.029	-	-
500 Hz (f)	0.007	-0.064	-0.109	0.381	-	-
1 kHz (f)	-0.001	-0.095	0.057	0.539	-	-
2 kHz (f)	0.026	-0.031	-0.139	0.142	-	-
4 kHz (f)	-0.015	-0.078	1.000	0.110	-	-
8 kHz (f)	-0.019	0.092	0.079	-0.057	-	-
0.5 Hz (t)	0.088	-0.174	-0.029	0.062	0.862	-0.231
1 Hz (t)	0.211	0.197	0.081	0.017	0.893	-0.344
2 Hz (t)	0.023	-0.024	-0.279	-0.016	0.887	-0.401
4 Hz (t)	0.434	0.044	0.482	0.413	0.916	-0.297
8 Hz (t)	0.155	-0.008	-0.202	-0.280	0.928	0.005
16 Hz (t)	0.051	-0.021	-0.041	-0.061	0.840	0.398
32 Hz (t)	0.132	0.108	-0.064	-0.150	0.803	0.518
64 Hz (t)					0.741	0.510

Table 3.8: *The factor coefficient matrix (matrix A) for the factor analysis (left) and the principal components analysis for the temporal domain (right) before Varimax rotation.*

The product of the factor coefficient matrix (matrix A) [4×14] (Table 3.7) and matrix B results in a new matrix [1×4] representing the new signal in the reduced subspace for the combined analysis. For the temporal analysis, the outcomes of each parameter is also normalized (Table 3.7), which results in matrix B [1×8]. The product of the factor coefficient matrix (matrix A) [2×8] (Table 3.8) and matrix B results in a new matrix [1×2] representing the new signal in the reduced subspace for the temporal analysis.

For example, consider the the continuous male speech-shaped signal (n39) from this database. Firstly, the spectra and modulation spectra of the signal are normalized, using Table 3.7. Then the result-matrix (C) is multiplied by the factor coefficient matrix (A), according to Equation 3.1 and the resulting values shown in Table 3.10 are obtained.

band	spectral	temporal		norm.-spectral	norm-temporal
1	67.31	0.0439		0.545	-1.042
2	69.41	0.0734		0.654	-0.968
3	69.79	0.0991	normalization	0.952	-0.955
4	62.91	0.1386	according to	-0.173	-0.961
5	60.50	0.1853	$\frac{band-\mu}{\sigma}$	-0.275	-0.935
6	56.41	0.2569		-0.027	-0.530
7	48.97	0.3370		0.082	-0.676

Table 3.9: *Normalization of the 14 parameters (Matrix C)*

$$\text{factor}_i = \sum_{1}^{14}(A_i * C_i) \tag{3.1}$$

	Original	Calculated	% Deviation
factor 1	-0.99	-0.99	0.55
factor 2	0.56	0.55	1.81
factor 3	0.21	0.18	14.42
factor 4	0.16	0.14	13.38

Table 3.10: *Calculated loadings compared with the original loadings. The deviation is merely due to rounding errors*

Table 3.10 shows that the deviation is, especially for the first two components, rather small (0.5%), and occurs purely due to rounding errors.

In order to determine which cluster is nearest to a new signal, the location of all clusters should be available. Table 3.11 gives the location of the 15 clusters for the factor analysis. Table 3.12 presents the location of the 10 clusters for the Principal Components Analysis from Appendix A.

Cluster	Situation	Dim1	Dim2	Dim3	Dim4
A	Industry	-1.072	-0.541	1.277	-0.116
B	Restaurant + ♀talker	-0.068	-0.535	0.319	0.323
C	Bandsaw	0.258	-1.737	-0.309	-3.867
D	Applause	-0.750	-2.069	-1.032	-0.170
E	Building site	1.768	-0.337	0.442	0.934
F	Traffic in car	-0.519	1.341	-1.740	-0.874
G	Quiet highway	5.096	-0.173	-0.858	-0.748
H	Football stadium	-0.232	-1.476	-0.517	0.669
I	♀talker	2.167	0.714	-0.743	0.501
J	Crickets	-0.624	-0.966	1.015	-2.929
K	Machine gun	1.798	1.773	-3.232	-2.565
L	Talking office	3.327	0.055	1.520	0.822
M	Thunderstorm	4.566	0.905	0.753	0.519
N	Jungle day	0.701	-2.943	-0.194	-1.641
O	Swamp night	1.369	-5.188	-1.956	-1.236

Table 3.11: *Location of the clusters (n=15) in the four dimensional domain obtained from the factor analysis.*

Cluster name	signal	dim 1	dim 2
A	Hans Boer	-1.272	-0.047
B	New tram & speech babble	-0.346	0.232
C	Building site	2.477	-0.589
D	Quiet highway	1.767	4.710
E	Babble (Engl.) low SNR	1.765	-1.089
F	Bar	1.143	-0.025
G	Bureau office	2.914	-0.064
H	Machine gun	-0.617	4.752
I	Talking office	3.915	-0.407
J	Crickets	-1.499	1.437

Table 3.12: *the location of the clusters (n=10) in the two dimensional domain of the principal components analysis (Appendix A)*

3.8 Appendix C - Deviating behavior of speech

One of the motivations for this study was to see whether noise and speech can be segregated according to differences in their spectro-temporal characteristics. Cumulative distributions are determined for each dimension as a function of their loading on that dimension for three different situations (see Figure 3.8):

1. signals consisting entirely of speech (n=10: squares).

2. signals consisting entirely or partly of speech or music including speech-shaped noise (n=34; triangles). All signals that contain speech or are created to approach the speech-signal.

3. Noise signals. all noises not presented in categories 1 or 2. (n=110: pluses).

If two cumulative distributions overlap, it is difficult to distinguish between the two situations represented. Figure 3.8 shows that noise and speech have similar loadings on the first two spectral dimensions. This indicates that it is difficult to distinguish between speech and noise using these dimensions. However, speech signals have higher loadings than noises on the dimension representing temporal behavior (dimension 1) and the spectral dimension representing the bandpass at 4 kHz (dimension 4). Since the percentage of variance explained by the temporal behavior is much larger than the percentage explained for the spectral dimension. temporal behavior is more suitable to discriminate between speech and noise. In other words, speech and noise are more distinguishable based on their temporal characteristic than on their spectral characteristics.

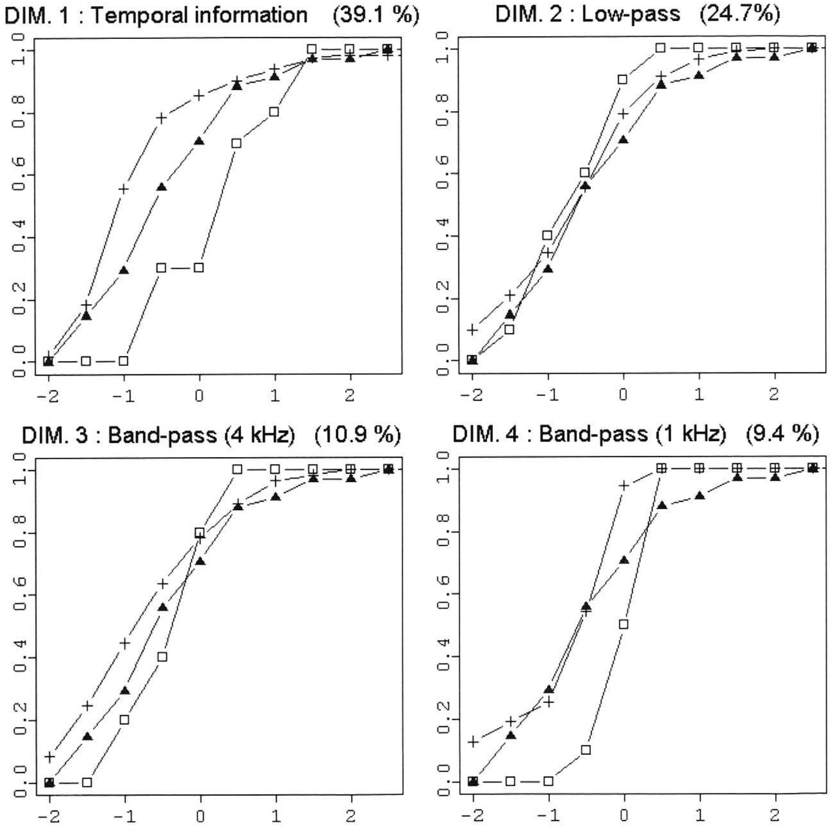


Figure 3.8: *Cumulative distribution of the proportion of signals (pluses), speech music and speech-shaped noise (triangles), and speech (squares) having a lower loading than given by the abscissa.*

3.9 Appendix D - Speech intelligibility and acoustical properties

Speech intelligibility can be restricted by acoustical parameters such as background noise and the limitations of spectral and temporal resolution. Recently, a database of noises has been developed in a similar way as described in this chapter. This section illustrates the facilities provided by this type of database in terms of speech intelligibility[§]. Speech intelligibility was assessed using two different procedures, both using the VU98 sentences (Versfeld et al., 2000):

1. by method of reproduction of unknown sentences according to the standard SRT test (Plomp and Mimpen, 1979).

2. by method of adjustment, according to the JFC procedure[¶] in which subjects have to adjust the level of noise until they can just follow a set of known sentences.

Due to time restrictions, fewer noises could be used for the SRT experiments (12) than for the JFC tests (43). Since the JFC and SRT gave comparable results[‖], results of the JFC were used for further analysis.

Following the PCA, results were similar to the results given by Fig. 3.4[**]. The contribution of each factor was determined using multiple regression with the JFC-outcome value as the dependent variable and the coordinates of the noises in the 3-dimensional plane as the explaining variables. The results are given in Table 3.13. The high loadings on the low-pass spectral component indicate that the relatively high energy in the low frequency bands (< 1kHz) is detrimental to speech intelligibility, resulting in higher SRTs. The difference in the correlation coefficients

[§] The temporal analysis differed from that described in section 3.2.2. The envelopes of the individual octave bands were determined, rather than the envelope of the broadband spectrum. For instance, the activity of 0.5 Hz is now determined by the average of the 0.5 Hz components over different octave carrier bands.

[¶] The JFC procedure (Just Follow Conversation) is a procedure in which the subject is asked to adjust the signal to noise ratio until a speech signal, consisting of concatenated known sentences, can be followed in the presence of a background noise (Larsby and Arlinger, 1994).

[‖] Since the JFC depends on the criterium given by the subjects. Individual correlation coefficients were determined between the JFC and SRT results for 12 noises. Most correlation coefficients were close to 1, indicating a similar trend for both measurements as a function of the noise type.

[**] The PCA resulted in three factors instead of four. The third factor was build up as if containing the third and fourth factor and could be considered as a high-pass filter.

for the high frequency factor may be attributed to audibility. Since most hearing-impaired subjects had high-frequency hearing losses and noises were scaled to equal dBA, increased energy for high frequencies was less disruptive to intelligibility than energy for the low frequencies. Furthermore, the negative correlation coefficient for the temporal dimension indicates that normal hearing subjects benefit from the gaps in the noise, whereas hearing-impaired subjects do not. This effect is denoted as masking release for speech throughout this thesis.

Group	temporal	low-pass	high-pass
NH	-0.43	0.71	0.18
HI	0.00	0.55	-0.16

Table 3.13: *Coefficients for multiple regression with JFC-outcome values as dependent variables and the components extracted as explaining factors. All correlation coefficients are significant* p*-value* < 0.01*, except for 0.00 (temporal HI)*

These results illustrate the importance of understanding the perception of amplitude modulated signals, and its' effects this may have on speech perception.

Chapter 4

Introduction to the perception of Sinusoidally Amplitude Modulated (SAM) signals

As has been discussed in the first three chapters of this thesis, important sounds in our natural environment, such as speech and music, are dynamic or time-varying. This means that they can be characterized by changes in amplitude or frequency, usually both. Hence, the ability to resolve dynamic changes must be important for the auditory processing of complex sounds.

4.1 Temporal resolution

Temporal resolution refers to the ability of a subject to detect changes over time within a signal and might be derived from one of the following measures:

1. **Sensitivity to amplitude modulation;** the sensitivity to temporal fluctuations, known as amplitude modulation, imposed on a signal.

2. **Gap-detection thresholds;** the duration necessary to detect a brief gap in a signal.

3. **Non-simultaneous masking;** the reduced sensitivity to a target due to a masker that preceded or followed the target. This is known as forward and backward masking, respectively.

This thesis concentrates on the sensitivity to sinusoidally amplitude modulated (SAM) signals.

A major difficulty in measuring temporal resolution in the auditory system is that changes in the time pattern of a sound, often also induce a change in the frequency spectrum of the signal. Additional cues are presented, which may aid detection. Hence the sensitivity to spectral changes rather than temporal resolution may be measured.

4.1.1 Inherent fluctuations and temporal processing

Another difficulty in measuring temporal resolution is that noise is defined by random fluctuations in the magnitude spectrum of the signal. These fluctuations may interfere with the temporal task. Figure 4.1 (panel A) shows two bands of Gaussian distributed noise of equal energy, a narrowband noise (left hand side) and a wideband noise (right hand side). The slow fluctuations in the narrowband signal are more pronounced than in the wideband signal. Panel B presents the power spectrum of the Gaussian bandpass noises. The spectrum of the squared envelope (panel C) of these noises is given by the DC-component and a sloping spectrum up to the bandwidth of the carrier (Lawson and Uhlenbeck, 1950). The area covered by the squared envelope spectrum depends on the bandwidth and the level of the carrier.

For noise carriers, the following rules are valid :

1. the ratio between the DC-component and the area covered by the modulation spectrum is constant as a function of level and bandwidth.

2. the bandwidth of the stimulus determines the range covered by the spectrum of the envelope.

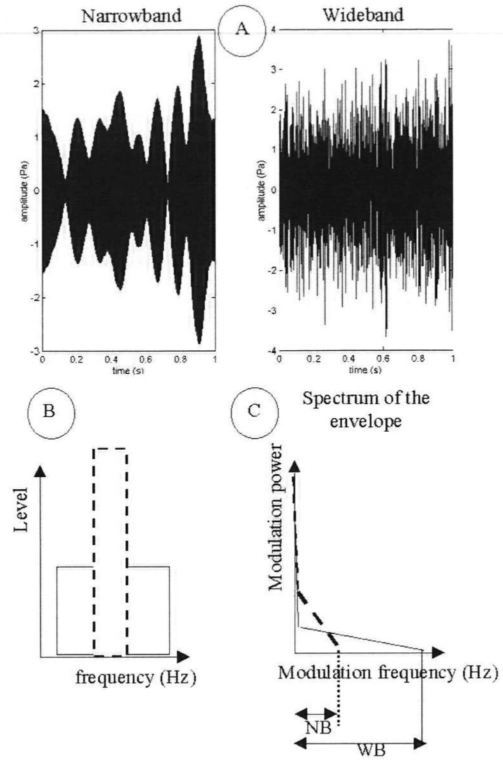


Figure 4.1: *Inherent fluctuations of a signal are determined by the bandwidth of the signal. For a narrowband signal (10 Hz wide; 707 Hz; left hand side panel A) these fluctuations are more prominent than for a wideband signal (1 octave; 707 Hz; right hand side). Panel B shows the power spectra of both signals. Panel C shows the spectrum of the envelope. The dashed and solid curves represent the narrowband and wideband signals, respectively*

Increasing the bandwidth, increases the the range of modulation frequencies and since the total area underneath the modulation spectrum is constant, the spectrum becomes broader and flatter (see panel C). Hence, widening the bandwidth, results in stronger fast fluctuations and weaker slow fluctuations. For a more extensive summary, see Dau (1996).

4.1.2 Modulation detection

The modulation detection threshold refers to the modulation depth required to detect the modulation of a carrier. Viemeister (1979) measured the ability to detect the modulation in amplitude of a broadband noise carrier as a function of the rapidity of the changes, or modulation frequency. The function relating the sensitivity to SAM (sinusoidal amplitude modulation) and the modulation frequency was termed as the temporal modulation transfer function (TMTF). The sensitivity to SAM is likely to depend on the bandwidth of the carrier. Due to the amount of inherent fluctuations in the unmodulated signal, the task to detect a modulated signal shifts from modulation detection to modulation discrimination. For narrowband carriers the effects will be largest for low modulation rates, whereas the detection of fast fluctuations will be less hampered. For larger bandwidths, the opposite is likely to occur. The sensitivity to high modulation rates will be affected, whereas the detection for low modulation rates becomes easier.

4.1.3 Gap-detection

Since abrupt gating does not have an effect on the long-term magnitude spectrum of a broadband white noise, the threshold for detecting a gap in a broadband noise provides a simple measure of temporal resolution. Gap detection thresholds for broadband noises are typically in the order of 2-3 ms (Plomp, 1964). Gap detection thresholds are relatively constant as a function of exposure level.

Slow fluctuations of noises may reduce the sensitivity to temporal gaps, by changing the gap-detection task into a gap-discrimination task. Such confusions with the actual gap are more likely to occur in the presence of slow inherent fluctuations than in the presence of fast fluctuations. Since slow inherent fluctuations are more prominent in a narrowband than in a broadband signal, gap-detection thresholds decrease as bandwidth increases (Shailer and Moore, 1983, 1985). If the relative bandwidth* is kept constant, thereby increasing the overall bandwidth as a function of frequency, gap detection thresholds decrease monotonically as center frequency increases, whereas gap detection thresholds are almost constant as a function of frequency when the absolute bandwidth is kept constant† (Eddins et al., 1992).

*Relative bandwidth : bandwidth expressed as a proportion of the center frequency

†Absolute bandwidth : bandwidth in *Hz*

This indicates that the absolute bandwidth is indeed the determining factor of gap-detection thresholds. However, the bandwidth of the auditory filter increases for increasing center frequencies. When the noise bandwidth is larger than the auditory filter bandwidth, the fluctuations at the output of the auditory filter are slower than those at the input. Hence, gap-detection was expected to increase with center frequency. The fact that gap-detection thresholds for absolute bandwidths do not decrease as frequency increases suggests that information can be combined across different auditory filters. Exciting different filters may increase performance, since the inherent fluctuations are frequency dependent and therefore asynchronous, whereas the gap is synchronous for the different frequencies. Hence, comparing a number of filter outputs for a constant bandwidth may increase performance.

4.1.4 Non-simultaneous masking

The sensitivity to a short target decreases when target and masker are presented simultaneously. However, the sensitivity to a target may also deteriorate when it is preceded by, or followed by a masker. This is referred to as forward, or backward masking, respectively.

Forward masking is reduced according to an exponential function as the temporal separation between target and masker increases. Jesteadt et al. (1982) showed that the sensitivity to a target is given by its detection threshold in quiet when the temporal separation between target and masker is as large as 100-200 ms. In addition, masking increases proportionally with the masker level. The amount of forward masking increases with the duration of the masker up to at least 50-200 ms. The nature of forward masking is still not fully understood. It can be explained by a reduction in the sensitivity of recently stimulated cells, or by the neural activity evoked by the masker itself. Since cochlear implant users experience forward masking effects on a similar scale as normal hearing subjects, forward masking is not purely an effect of ringing on the basilar membrane (Shannon, 1990).

The concept underlying backward masking is also not fully understood. Highly trained subjects show hardly an effect of backward masking, indicating that it may be based on confusion. Duifhuis (1973) reported that monaural[‡] backward masking

[‡]Monaural : a situation where the signal reaches *one* ear

appears to be far more prominent than dichotic[§] backward masking, which indicates that the origin of backward masking is mainly located peripherally, but also partly central.

4.2 The perception of SAM-signals

4.2.1 Modulation detection for broadband carriers

The TMTF for broadband signals can be considered as a valid representation of temporal resolution, since the long-term frequency spectrum of a broadband sound does not provide additional cues to facilitate temporal processing. The TMTF for broadband carriers is independent of modulation frequency up to about 16 Hz. Increasing the modulation frequency beyond 16 Hz reduces the sensitivity to SAM by approximately 4-5 dB/oct. up to approximately 1000 Hz for which SAM cannot be detected (Viemeister, 1979). The shape of the TMTF can be seen as a low-pass filter with a cut-off frequency (-3 dB) at approximately 50 Hz. Up to 16 Hz, performance is mainly limited by the ability to detect differences in amplitude, for modulation frequencies above 16 Hz temporal resolution starts to play a role. The sensitivity to SAM, as determined by the TMTF is almost independent of level using a broadband noise carrier, except for exposure levels below 20-30 dB SL, where the sensitivity to SAM starts to decrease (Viemeister, 1979).

4.2.2 Modulation detection for pure tone carriers

Chapter 2 illustrates how the power spectrum of a modulated sinusoid is built up of three spectral components. These are the frequency of the carrier and sidebands on each side of the carrier frequency each separated from the carrier by the modulation frequency (see Figure 2.1). If the bandwidth of the auditory filter centered at the center frequency is sufficiently small with regard to the modulation frequency, the sidebands may form a cue on which modulation can be detected. Similar to the TMTF for broadband noises, the sensitivity to SAM for sinusoidal carriers remains constant as a function of modulation frequency at low modulation frequencies, and decreases by approximately 5-8 dB/octave for higher modulation

[§]Dichotic : a situation where the signal reaching the ears is *different* for both ears frequencies. However, the TMTF for pure tone carriers differs from the TMTF for broadband carriers in two ways. Firstly, the sensitivity to SAM using pure tone carriers increases for modulation frequencies larger than given modulation frequency (Kohlrausch et al., 2000). The modulation frequency, after which the sensitivity to SAM increases, depends on the carrier frequency. The sensitivity to SAM for broadband carriers continues to decrease. Secondly, the initial flat proportion of the TMTF for sinusoidal carriers extends to approximately 100-130 Hz. This is relatively high compared to the TMTF for broadband carriers, which extends to approximately 50 Hz. It is likely that the discrepancy between these two cut-off frequencies can be attributed to the difference in inherent fluctuations in the time-waveforms. The part of the TMTF for pure tone carriers at which the sensitivity to SAM decreases for modulation frequencies above 100-130 Hz, presumably presents the part for which temporal resolution is limited by the human ear. The third part of the TMTF, in which the sensitivity to SAM increases is likely to reflect the ability to detect spectral sidebands. For carrier frequencies below approximately 1 kHz, the sensitivity to SAM for modulation frequencies above 100-130 Hz increases rather than decreases. This is probably due to the relative narrow bandwidths of the auditory filters at these frequencies ($ERB_{1kHz} = 133$ Hz), offering the ability to detect the modulation based on the spectral sidebands for modulation rates above 130 Hz.

The sensitivity to SAM generally increases with carrier level. At very low levels, the reduction in sensitivity to SAM as a function of the modulation frequency is also reported for 1 kHz carriers. For these low sensation levels, the spectral sidebands are below the detection threshold, resulting in a reduced sensitivity to SAM.

4.2.3 Modulation detection as a function of the bandwidth

The inherent fluctuations of the carrier (see Fig. 4.1) may reduce the sensitivity to SAM by changing the modulation detection task into a modulation discrimination task, with the inherent fluctuations serving as a reference depth and the modulation of the implied envelope as the change in modulation depth.

Increasing the bandwidth in terms of inherent fluctuations, leads to:

1. a reduction in the modulation power of low modulation frequencies.

2. an increasing modulation power of high modulation frequencies.

Hence, decreasing bandwidth is likely to affect the sensitivity to SAM. In addition, spectral sidebands may influence the sensitivity to SAM.

Several studies have shown an increase in temporal resolution as center frequency increases using relatively wide bands of noise (Viemeister, 1979; Bacon and Viemeister, 1985; Formby and Muir, 1988). However, in most studies, the bandwidth increased with the center frequency of the carrier. Studies that measured the sensitivity to SAM as a function of carrier frequency whilst keeping the absolute bandwidth constant, found no effect of center frequency (Eddins, 1993; Dau et al., 1997a). Dau et al. (1999) measured the sensitivity to SAM as a function of bandwidth using three different modulation frequencies. The bandwidth ranged from 1 Hz to 6 kHz (using a fixed upper cut-off frequency of 6 kHz). Sensitivity to SAM was highest when using a carrier bandwidth of 1 Hz and reduces up to the bandwidth corresponding to 2-4 times the applied modulation frequency. The sensitivity to SAM increases with increasing carrier bandwidth for bandwidths wider than 2-4 times the modulation frequency.

Besides the bandwidth, the sensitivity to SAM depends on the type of noise. Dau et al. (1999) differentiated between three different types of noise: Gaussian noise; multiplied noise; and low-noise noise[¶]. The sensitivity to SAM is low for the Gaussian or multiplied noise carriers for low modulation frequencies and high for fast modulation frequencies This is in agreement with the noise characteristics of the envelope spectrum. The sensitivity to SAM increases relatively fast for the

[¶]These three noises differ in envelope and envelope spectrum from each other.

1. Envelope: the envelope of the low-noise noise is flat as a function of time relative to the other two noises.

2. The spectrum of the squared envelope for a :

   (a) Gaussian noise is given by the triangular shape of figure 4.1C

   (b) multiplied noise is, ideally, given by a rectangular shape ranging to half its modulation frequency

   (c) low noise noise is, ideally, given by a triangular shape with fewer slow fluctuations than fast fluctuation, with its maximum at the bandwidth of the carrier.

multiplied noise compared to the Gaussian noise for modulation frequencies larger than half the carrier bandwidth. In addition, the sensitivity to SAM using a low noise noise carrier is high for low modulation frequencies and relatively low for high modulation frequencies.

4.2.4 Stimulus duration

Most studies measured modulation detection using a constant stimulus duration. However, it seems plausible that the sensitivity to SAM is, at least to some extent, determined by the duration of the stimulus. In an extreme situation, modulation detection based on a single cycle would lead to poorer results than modulation detection based on multiple cycles. The sensitivity to SAM does generally increase with duration (Viemeister, 1979; Lee and Bacon, 1997). The concept of 'critical duration' is adopted, to describe the duration for which longer exposure times will not lead to a higher sensitivity to SAM for that modulation frequency. This duration generally corresponds to 4 or 5 cycles of the modulator (Lee and Bacon, 1997). The concept follows the multiple-looks model proposed by Viemeister and Wakefield (1991). This model suggests that signals are sampled at a fairly high rate (integrated over approximately 3 ms). These samples, or "looks", are stored in a short-term memory, which can be accessed easily and processed selectively.

4.2.5 Modulation discrimination

Modulation depth discrimination extends modulation detection by offering information on the ability of the auditory system to use, rather than detect, envelope fluctuations. The just noticeable difference in modulation depth for a broadband noise carrier with a reference modulation depth of less than -10 dB is smaller than, or equal to the sensitivity to a change in SAM for relatively low reference depths (-25 dB) when the difference in modulation power (20 $\log_{10}[m_c^2$ - $m_s^2]$) is plotted as a function of the standard modulation depth (20 $\log_{10}[m_s^2]$). For reference depths above -10 dB, the just noticeable differences in modulation depth increases rapidly (Wakefield and Viemeister, 1990). This behavior has been reported for modulation rates up to 400 Hz.

Weber's law (see Eq. 4.1) states that the smallest detectable change in a stimulus (ΔS), is proportional to the magnitude of that stimulus (S).

$$\frac{\Delta S}{S} = C \tag{4.1}$$

The constant (C) is referred to as the Weber fraction. For modulation discrimination, the function relating the just noticeable difference in modulation depth and the reference depth indicates that this law holds for modulation depths up to approximately -7 dB (m = 0.45) independent of carrier type (von Fleischer, 1980; Ozimek and Sek, 1988; Wakefield and Viemeister, 1990). The sensitivity to a change in SAM increases for reference depths larger than -7 dB.

4.3 The effects of hearing-impairment

4.3.1 Loudness recruitment

The first symptom of hearing-impairment is generally given by a reduced audibility of signals. However, the perception of supra-threshold signals can also be abnormal, for example, the perception of loudness. Loudness is the sensation that is attributed to a sound on a scale between the detection threshold and the uncomfortable loudness (UCL). Hence, loudness is a subjective measure and may depend on a variety of parameters‖. Whereas detection thresholds are elevated for hearing-impaired subjects, the UCL is generally similar to or lower than the UCL for normal hearing subjects (see dotted line Fig. 4.2). Most hearing-impaired subjects with hearing losses of cochlear origin** show this phenomenon, which is referred to as loudness recruitment (Steinberg and Gardner, 1937).

The perceived loudness for hearing-impaired subjects of signals presented at thresholds and at UCL is comparable to that of normal hearing subjects, namely just audible and too loud. This suggests that the growth in loudness as a function of intensity must be faster for hearing-impaired than for normal hearing subjects. This is illustrated by the difference in the slopes of the dotted and solid lines in Figure 4.2. New insights provided by Buus and Florentine (2002) suggest that the loudness

‖Besides level, loudness is known to depend upon bandwidth, frequency, duration and temporal structure (*e.g.* amplitude modulations).

**In the remainder of this thesis we will refer to hearing-impairment when we refer to hearing losses of cochlear origin or sensorineural origin, unless explicitly stated otherwise.

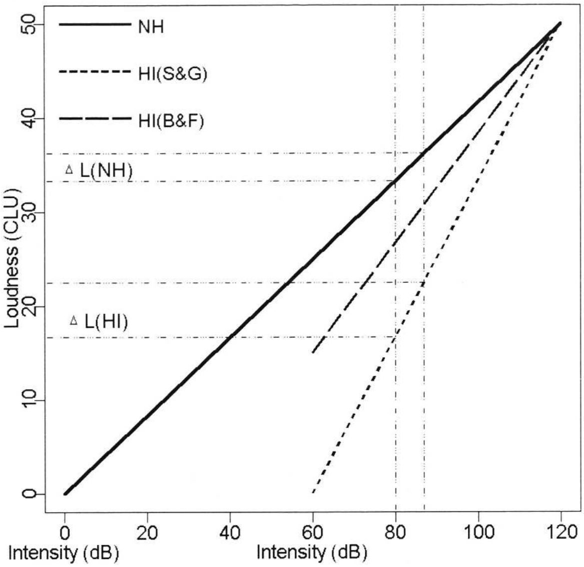


Figure 4.2: *Perceived loudness (ordinate) as a function of stimulus level (abscissa) for three different configurations. Normal hearing subjects (NH); a conventional view of loudness recruitment for hearing-impaired subjects (loudness equal to 0 at detection threshold) according to Steinberg and Gardner (1937) (S&G); loudness recruitment following new insights by Buus and Florentine (2002) (B&F)(loudness above 0 at detection threshold)*

at, or just above threshold may differ for normal hearing and hearing-impaired subjects. The loudness just above the threshold is thought to be higher than 0, which suggests that recruitment is partly attributable to *softness imperception*. Although the evidence is convincing, there are also data that cannot be explained by this hypothesis.

Loudness recruitment is often linked to an impaired cochlea and more specifically to an impairment of the inner and outer haircells on the basilar membrane. The basilar membrane of the normal hearing ear acts non-linearly. This implies that increasing the input by a certain factor does not result in a similar increment at the output of the basilar membrane (Robles and Rich, 1986). For low and high input levels (below 20-30 dB SPL and above 80-90 dB SPL), the input-output function for normal hearing ears acts more linearly. For intermediate levels, the slope is much more *compressive*, indicating that the output increases by a smaller amount than the input. This non-linearity mainly occurs when the stimulating frequency is close to the characteristic frequency of the neuron[††], the input-output relation becomes more linear for larger frequency separations. In the case of cochlear damage, the compressive part for medium levels becomes more linear.

4.3.2 Temporal resolution for hearing-impaired subjects

Glasberg et al. (1987) measured forward masking for unilaterally hearing-impaired subjects[‡‡]. Their results indicate that the rate of recovery from forward masking, serving as a measure of temporal resolution, occurs more rapidly for normal ears than for impaired ears when the comparison is made at equal SPL. However, comparison at equal SL, indicated that the differences between normal and hearing-impaired ears is much reduced. Since forward masking depends upon the exposure level, the slower rate of recovery may simply reflect low SLs. Alternatively, differences can be explained by assuming level-dependent compression on the basilar membrane (Plack and Oxenham, 1997). As mentioned above, the input-output relation of the basilar membrane is considered to be more linear for hearing-impaired listeners, whereas for normal hearing listeners at intermediate levels the input-output function is considered to be much more compressive. Hence, for a constant level difference in the input, the induced effect at the output is much larger for hearing-impaired subjects than for normal hearing subjects (see also Fig. 4.2). For low sensation levels, the input-output function of the basilar membrane for normal hearing subjects is less compressive and more comparable to hearing-impaired listeners. Hence, the induced differences in the output of the compressive non-linearity are more similar to each other. In other words, differences in forward masking, may also reflect a difference in the input-output function of the basilar membrane.

Hearing-Impairment and SAM-perception

Bacon and Gleitman (1992) showed that the sensitivity to SAM for hearing-impaired subjects with relatively flat losses is higher when exposure occurred at equal, low,

[††] neurons are tuned to a certain characteristic frequency, which means that the response of this neuron is largest when exposed to this frequency. However, exposing to other frequencies will also cause excitation in a neuron tuned to a distant frequency.

[‡‡] Unilateral hearing-impaired subjects : subjects with a one-sided hearing loss; one normal ear and one hearing-impaired ear.

sensation levels. However, the sensitivity to SAM using a broadband noise carrier, is lower for hearing-impaired subjects with sloping losses than for normal hearing subjects (Bacon and Viemeister, 1985). This may partly be attributed to the inaudibility of high frequencies for the hearing-impaired listener. Normal hearing subjects, using band-limited noise, show a reduced sensitivity to SAM that was similar to hearing-impaired subjects (Bacon and Viemeister, 1985). The TMTF as a function of bandwidth shows no clear interaction between impairment and bandwidth as long as stimuli are presented at at least 25-30 dB SL (Hall et al., 1998). In addition, the study by Formby and Muir (1988) indicates that the sensitivity to SAM is higher for a high carrier frequency band than for low carrier frequency bands. Hearing-impaired subjects are generally less sensitive to these higher frequencies than normal hearing subjects, which may explain the reduced sensitivity to SAM*.

Cochlear impairment may affect modulation perception not only in terms of the audibility of the signal, but also as a result of the lack of compression for hearing-impaired subjects. This lack of compression may influence the intensity relations within a signal by increasing the difference in loudness between the peaks and valleys for hearing-impaired subjects with regard to normal hearing subjects. Figure 4.2 shows the difference in loudness for a constant difference in intensity, as, for instance, induced by a modulated carrier. This difference in loudness is clearly larger for hearing-impaired subjects than for normal hearing subjects and may result in a better performance. Expressed in terms of the RMS†, unmodulated signals are comparable in loudness to modulated signals (Moore et al., 1999). Other matching experiments with the modulation depth as the adjustment parameter using unilateral hearing-impaired subjects are described by Wojtczak (1996) and Moore et al. (1996). In order to prevent loudness from acting as a cue, the unmodulated carriers are equalized in loudness between both ears. A given modulation depth in the impaired ear is matched with a larger modulation depth in the normal ear. For modulation rates above 8 Hz, the differences in modulation depth for the normal ear

*Although the low-pass signals were effectively wider than the high-pass signals, signals were high-pass filtered by means of the filtering characteristics of the headphones. Since no actual high-pass filtering took place, the inherent fluctuations of the high-pass signal may actually be weaker than the presumed 2.5 kHz bandwidth. Given the importance of these inherent fluctuation to modulation detection, the outcome should be interpreted with care.

†RMS: Root-Mean-Square, a quantity obtained by squaring the instantaneous value of the waveform, taking the average of the squared value over time, and taking the square root of this average.

and the impaired ear declined as modulation depth decreased. Apparently, loudness recruitment affects the perception of dynamic aspects of sounds by enlarging the perceived modulations. It is likely that these effects are caused by the lack of fast compression, which results in an increased growth of loudness. This increased growth of loudness increases the differences in perceived loudness for the peaks and valleys, as suggested in Figure 4.2.

4.3.3 Impaired spectral resolution

Cochlear damage often results in a reduced frequency selectivity. Frequency selectivity refers to the ability of a subject to distinguish between two signals according to their spectral differences.

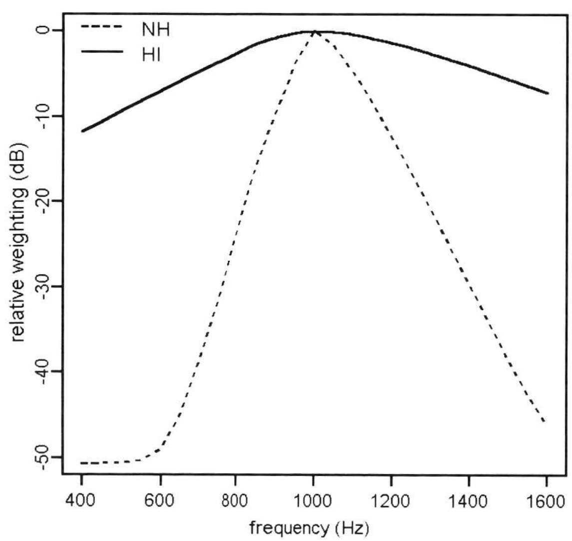


Figure 4.3: *Auditory filter for a normal hearing subject (dashed line) and a hearing-impaired subject (solid line)*

It is widely accepted that the basilar membrane can be considered as a bank of overlapping filters (Fletcher, 1940). Thus any complex sound, such as speech, is subjected to filtering and is split into certain components within a certain frequency range. These components are coded independently in the auditory nerve, provided that the spectral differences in these components are sufficiently large. Most methods for estimating the auditory filter shapes at a specific frequency are based on masking. Figure 4.3 shows two examples of auditory filter-shapes‡. First consider the filter shape for normal hearing subjects (dashed line). The data are obtained by a method described by Moore and Glasberg (1987), based on the audibility of a pure tone, placed symmetrically or asymmetrically in the spectral notch of a masker. The ability to detect a target in the presence of these notched noises, is determined by the energy of the masker falling within the auditory filter tuned to the signal frequency§. An alternative method for describing the bandwidth of the auditory filter is given by the equivalent rectangular bandwidth (ERB). The ERB corresponds to the bandwidth of a rectangular filter that has the same peak transmission as the original auditory filter and has the same intensity transmission as the original filter for a white noise.

Auditory filters broaden as frequency increases. For hearing-impaired subjects and for normal hearing subjects at higher exposure levels, filters become asymmetric. Auditory filters of hearing-impaired subjects are generally broader than the filters of normal hearing subjects and, on average, broadening increases with hearing loss, mainly due to a more shallow low-frequency skirt (p_l) (Glasberg et al., 1987; Leeuw and Dreschler, 1994). These broader filters result in a reduced spectral selectivity.

4.4 Co-modulation Masking Release (CMR)

According to the power spectrum model for masking, the audibility of a signal presented in a background noise is determined by the signal-to-noise ratio at the output of the auditory filter centered at, or close to, the signal frequency. In some

‡Data adapted from Franck et al. (2004), with permission of the Author

§The filter shape (see Fig 4.3) is determined by the roex formula:

$$\begin{aligned} W(g) &= (1-r)(+p_l g)e^{-p_l g} + r \qquad \text{if} \quad f < f_c \\ &= (1-r)(+p_u g)e^{-p_u g} + r \qquad \text{if} \quad f > f_c \end{aligned} \tag{4.2}$$

where: $g = \frac{|f-f_c|}{f_c}$

in which the attenuation (ordinate) $W(g)$ is expressed as a function of the normalized frequency g based on three parameters: slope of low-frequency flank (p_l), the high-frequency flank (p_u) and the dynamic range of the filter r. Constants p_l and p_u determine the slopes of the filterskirts and, hence, the overall bandwidth. When the two filters, given in Figure 4.3 by roex$_{NH}$(38.1, 22.9, 8.7*10^{-6}) and roex$_{HI}$(7.6, 5.2, 0.0083) are compared, it is obvious that higher p-values are found for normal hearing subjects corresponding to steeper and thereby narrower filters.

cases, the auditory system uses information provided outside the auditory filter in order to improve performance. In these cases, the power spectrum model of masking fails. One of these cases is CMR. CMR refers to the increased sensitivity to a target, usually a pure tone, due to information provided on the fluctuations of the envelope of the masker, outside the filter tuned to the signal-frequency. This can occur in two ways :

1. increasing the bandwidth of the masker beyond the critical bandwidth (Zwicker, 1961; top Figure 4.4).

2. adding a second band of noise on a spectrally remote distance (off frequency band) having the same temporal envelope as the masker masking the target, also referred to as the on frequency band (bottom Figure 4.4)

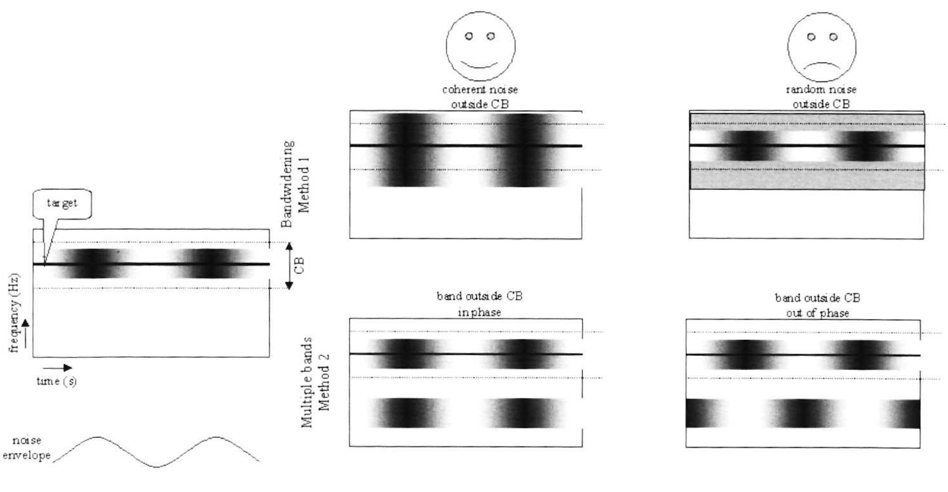


Figure 4.4: *Two methods leading to an increased sensitivity to a masked target (CMR). Increasing the masker bandwidth (top) beyond the critical bandwidth (CB) leads to a higher sensitivity to a target when the noise is modulated by the same modulation pattern (coherent) as the noise inside the CB (left hand side) and an unaltered sensitivity when random noise is added, according to the concept of the critical band (right hand side). Adding a second band (bottom), increases the sensitivity to a target when this band has the same envelope as the first band masking the target (left hand side). Adding an envelope with a different phase than the band masking the target does not change the sensitivity to a target.*

4.4.1 CMR of the first kind: band widening

Hall et al. (1984) measured signal detection in the presence of bandlimited noise. In random noise, containing irregular fluctuations in amplitude independently distributed over different frequency regions, signal detection thresholds increased as the bandwidth increased up to the bandwidth of the critical band. Increasing the bandwidth beyond the critical band, did not result in sensitivity to the target being further reduced, as predicted by the power spectrum model of masking and the concept of the critical bandwidth (Zwicker, 1961). In a fluctuating noise, modulated by a low-pass filtered noise at 50 Hz and thereby comodulated across critical bands, thresholds *decreased* when the noise bandwidth was increased beyond the critical bandwidth. Besides the phenomenon, CMR has been adopted as the measure of the benefit due to the coherence of the envelopes. The sensitivity to a signal presented in noise can be elevated by as much as 15 dB (Hall and Grose, 1988), mainly attributable to the across channel effects. Moore et al. (1993) mentioned that there was even a significant CMR effect for bandwidths smaller than the ERB of the auditory filter. This implies that CMR is at least partly due to within channel cues and is not purely a result of comparing different filter-outputs to each other.

4.4.2 CMR of the second kind: multiple bands

Hall et al. (1984) showed that a similar effect of CMR can be obtained by adding a second band of noise outside the critical band. The sensitivity to a target increased when this co-modulated off-frequency band was added. Increasing the number of co-modulated noise-bands increased the CMR effect up to 16 dB (Hall et al., 1990). Adding noise bands below the signal frequency resulted in a similar CMR as adding bands above the signal frequency. Noise bands closer to the target band resulted in more CMR than those at a larger spectral distance (Hall et al., 1990). The precise origin of this proximity effect is not entirely clear and may also reflect the contribution of within channel processing.

The magnitude of CMR for amplitude modulated noise is largest when the masker and off frequency bands are presented in phase (0^0) and CMR may even be reduced to 0 dB when the off-frequency bands are modulated 180^0 out of phase (Fantini, 1991). Amplitude modulated noise bands showed a CMR-effect of 2 dB when presented 180^0 out of phase, whereas the in-phase condition resulted in a CMR effect of 20 dB (Eddins and Wright, 1994). Hall et al. (1990) showed that the CMR-effect was reduced by adding noise bands, which were modulated out of phase with the on-frequency band.

Dichotic CMR has also been reported. Since CMR is similar for monaural and dichotic presentation, peripheral interactions are apparently not critical to CMR (Moore and Emmerich, 1990; van de Par and Kohlrausch, 1998). Just as for monaural CMR, dichotic CMR increases as more bands are added. Additionally, more CMR has been reported as maskers are closer to the target frequency (Hall et al., 1990).

4.4.3 CMR for hearing-impaired subjects

Hall and Grose (1988, 1989) measured CMR by broadening the bandwidth of the masker for normal hearing and hearing-impaired subjects. Both groups show a substantial CMR effect. Some hearing-impaired subjects show normal or near-normal results. However, most hearing-impaired subjects experience reduced CMR with regard to normal hearing subjects. It is most likely, that this is due to the broader filters for hearing-impaired subjects or to the low sensation level. Subjects with very broad filters show little, if any, CMR (Moore et al., 1993; Grose and Hall, 1996). CMR is largest for experiments carried out at the higher end of the range of comfortable levels (Hall and Grose, 1994). When experiments are carried out at equal SLs, CMR is comparable for normal hearing and hearing-impaired subjects. Furthermore, hearing-impaired subjects show a significantly reduced CMR compared to normal hearing subjects when bands with a different fluctuation pattern as the on-frequency band are added. This may reflect that hearing-impaired subjects experience increased difficulties in separating different sources based on common modulation.

4.5 Modulation detection or discrimination interference (MDI)

The sensitivity to SAM of a target may be reduced due to the presence of other modulated signals that are presented simultaneously. This may occur for broadband signals, with masker and target having partially overlapping spectra but also for two clearly distinctive carriers. In the first case, the reduction in sensitivity to SAM is referred to as modulation masking (Houtgast, 1989; Bacon and Grantham, 1989). In the second case, it is referred to as modulation detection interference or modulation discrimination interference, both abbreviated as MDI (Yost and Sheft, 1989). Typically, MDI is measured using three different conditions:

1. Probe Alone (PA); the sensitivity to SAM of a carrier.
2. UNModulated maskers (UNM); the sensitivity to SAM of a carrier in the presence of an *unmodulated* carrier.
3. MODulated maskers (MOD); the sensitivity to SAM of a carrier in the presence of a *modulated* carrier.

MDI is also used as the measure of the reduction in sensitivity to SAM, in dB, as a result of added modulated maskers, using the formula:

$$MDI = 20\log_{10}(m_{mod}) - 20\log_{10}(m_{unm}) \tag{4.3}$$

in which m reflects the modulation detection threshold in the task given by the subscript. MDI can grow as large as 20 dB (Yost and Sheft, 1989) and depends on factors such as the spectral location of masker and target. It has been suggested that MDI occurs because target and masker modulations are processed together in a modulation filter bank, where each filter is tuned to a different modulation rate (Yost and Sheft, 1989; Yost et al., 1989). Most studies however suggest that MDI results, at least partly, from perceptual grouping (Yost and Sheft, 1989; Yost et al., 1989; Hall and Grose, 1991; Moore and Shailer, 1992)[¶], target and masker modulation are fused into a single percept, reducing the ability to detect the target modulation.

¶ perceptual grouping refers to the process by which different components are grouped into one auditory object(see also 4.5.2)

4.5.1 Within channel or across channel processing

The effect of a 4 kHz masker on a 1 kHz target (Yost and Sheft, 1989), clearly indicates that MDI can be considered as a central process (across-channel processing). However, Moore and Shailer (1992) exposed some effects indicating that MDI is also due to a more peripheral process[||] (within channel processing). For example, the sensitivity to SAM decreases with frequency separation for the carriers in the MOD and UNM task. This indicates that the presence of the energy from the masker may have reduced the sensitivity to SAM of the target, as a result of the spread of excitation. Adding a background noise to mask possible within-channel cues, removed the proximity effect.

In addition, the reduction in sensitivity to SAM is larger when the masker carrier is higher in frequency than the carrier of the target, which indicates that modulation detection mainly occurs on the high-frequency flank of the excitation pattern (Bacon, 1999). The nonlinear growth of excitation may enhance the changes evoked by SAM on the high-frequency side (Glasberg and Moore, 1990; Moore and Shailer, 1994). Other studies (Bacon and Moore, 1993; Yost and Sheft, 1994) indicated that MDI is almost unaffected by the frequency separation of the target and masker. MDI is similar in size for a masker carrier that is higher or lower in frequency than the target. Adding two maskers results in an increased MDI, however adding additional maskers barely affects the sensitivity to SAM (Bacon and Moore, 1993). Dichotic MDI indicates that MDI is both an across channel effect and an effect that may occur across ears (Yost and Sheft, 1990; Bacon and Opie, 1994; Mendoza et al., 1995b).

4.5.2 MDI and perceptual grouping

Perceptual grouping has been mentioned as one of the effects contributing to MDI (*e.g.* Yost and Sheft, 1989; Yost et al., 1989; Hall and Grose, 1991; Moore and Shailer, 1992). Perceptual grouping refers to the process whereby different sound elements are assigned to a single auditory object. The concept is well known from the

[||] MDI is often subdivided into two components: carrier specific MDI and modulation specific MDI. Carrier specific MDI may occur due to the difficulties in hearing out the target frequency. Modulation specific MDI may occur due to the difficulties in distinguishing the modulation of the target from the modulation of the maskers.

visual system**, of which an example is given in Figure 4.5. The effect of perceptual grouping was introduced in auditory research by Bregman (1994).

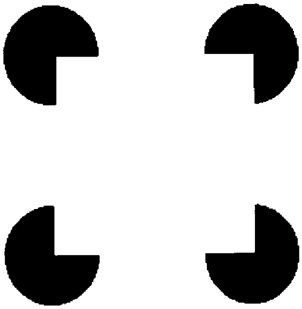


Figure 4.5: *An example of perceptual grouping: the four incomplete circles together form a white square without the actual contours of the square. In other words, the missing parts of the four circles are grouped into a square.*

The most important cues that may lead to perceptual grouping are:

1. **common onsets:** this cue can be attributed to the fact that most physical systems, when excited, produce complex sounds whose components have the same event timing. Although less prominent, the offset cue is also considered as a cue.

   The effect of MDI is largest for synchronous presentation and is greatly reduced when the masker was gated on before (and gated off after) the target (Hall and Grose, 1991). The reduction in MDI due to asynchronous gating has been confirmed in many studies, although all have found some residual MDI, even when the masker is presented continuously (Moore and Shailer, 1992; Mendoza et al., 1995a,b). In addition, maximum MDI does not always occur for similar gating windows (Shailer and Moore, 1993).

2. **common frequency or amplitude modulation**††**:** common frequency or amplitude modulation of individual components is a powerful cue stimulating perceptual grouping‡‡.

   MDI (Moore et al., 1991; Bacon and Konrad, 1993) and modulation masking (Houtgast, 1989) show a clear effect of tuning for modulation frequencies, with most masking occurring for identical modulation rates and a reduced masking effect for increasing differences in modulation frequency for target

**introduced by Wertheimer in 1912 and referred to as gestalt "put together[German]" school

††also referred to as common fate.

‡‡In some cases, common onsets is regarded as a special case of common amplitude modulation (a single modulation period). However, within this thesis both cues are considered as separate cues.

and masker. Grantham and Bacon (1991) reported a similar tuning effect for dichotic measurements. Analogous to the carrier frequency domain, a critical bandwidth concept has been described, indicating that the sensitivity to SAM reduces as long as the bandwidth of the masker modulation is increased. For modulation bandwidths broader than approximately 0.5 octave, the sensitivity to SAM remains constant (Houtgast, 1989).

These results are supported by Bregman et al. (1985, 1990). They demonstrated that two components that are modulated coherently, are more likely to be heard as one sound than when two components are modulated at different rates. However, grouping as a function of modulation rate was clearly reduced for small changes in modulation rate, whereas MDI is only influenced by large differences in modulation rate for target and masker. In other words, perceptual grouping is much more sharply tuned in the modulation frequency domain than is shown by MDI and modulation masking experiments.

3. **harmonic relationship:** components that are harmonically related to a common fundamental frequency tend to fuse together. In addition, harmonic pairs of complexes that are closer in frequency are more likely to be fused.

   There was little systematic effect of the number of maskers and of the harmonic relations for the carriers of these maskers, for masking produced by modulated and unmodulated maskers on the sensitivity to SAM (Bacon and Moore, 1993; Moore and Shailer, 1994).

4. **rhythm:** a rapid sequence of tones may be perceived in two ways. As a single perceptual stream or as if signals were coming from two or more sources. The rhythm of alternation between two signals will determine wether grouping or segregation occurs.

   MDI is barely affected by phase relations of the modulators of the target and the masker (Yost and Sheft, 1989; Bacon and Konrad, 1993). The lack of phase effect is difficult to reconcile with an explanation in terms of perceptual grouping. Carriers, that are amplitude modulated or frequency modulated in phase, tend to be perceived as more strongly fused than carriers that are modulated with different phases (Bregman et al., 1985; Moore and Emmerich, 1990). Thus, if perceptual grouping is responsible for MDI, more masking is to be expected when the modulation of the masker is in phase with the modulation of the target relative to a masker that is out of phase (Moore et al., 1991). As a special case, the masking effect of a random amplitude modulated envelope (RAM)* on a sinusoidal amplitude modulation is determined. MDI caused by RAM maskers on a SAM target is comparable to MDI for SAM maskers on a SAM target (Mendoza et al., 1995a).

5. **higher level cues:** several cues, which promote perceptual grouping, such as spatial cues and cues based on timing communalities, have already been discussed.

   Preceding the MDI task by a 'cue' tone, to identify the carrier of the target, had little influence on MDI (Moore and Shailer, 1992). Introducing 'cue' tones, to provide a cue about the leading target, reduced MDI. In addition, introducing modulated maskers as 'cue' tones indicated no residual MDI (Oxenham and Dau, 2001). Apparently, MDI occurs based on the inability to separate the target modulation from the masker modulation due to a grouping of both carriers.

Based on these findings, perceptual grouping may be assumed to be an important prerequisite for MDI. However, it would be expected that some aspects would be stronger in perceptual grouping than have been reported for MDI. Examples are tuning of amplitude modulation frequencies in the MDI task and the lack of phase-dependency of the envelopes for MDI.

4.5.3 MDI and hearing-impairment

Little research has focussed on differences between normal hearing and hearing-impaired subjects in MDI. The performance of listeners with cochlear hearing loss on MDI tasks appears to depend upon the spectral configuration of the hearing loss. Grose and Hall (1994) reported a similar MDI for normal hearing listeners and listeners with relatively flat hearing losses. Bacon and Opie (2002) observed an increased MDI for hearing-impaired listeners with sloping pure tone audiograms, with largest losses for high frequencies, when the target carrier is presented in the region of hearing loss relative to normal hearing subjects. Hence, the difference in

*using a small band noise as a modulator instead of a pure sinusoid for SAM

MDI may also reflect the effect of sensation level†. Bacon and Opie (2002) indicated that differences between the normal ear and impaired ear are generally small for unilateral hearing-impaired subjects. Measuring masking patterns in the modulation domain using modulation masking indicates that modulation filters appeared to be wider for hearing-impaired subjects than for normal hearing subjects (Lorenzi et al., 1997).

4.6 MDI and CMR

Although CMR and MDI are quite different from each other, both effects come forward in the presence of modulated envelopes. CMR relies on the detection of a signal in a modulated background, by popping out due to coherence in maskers across frequency. MDI relies on the inability to hear out the target modulation in the presence of modulated maskers. This section focusses on how CMR and MDI may be related.

The most obvious coherence in the envelope is obtained by SAM-signals. CMR increases as the modulation depth of the off-frequency band increases (Fantini, 1991). As the modulation depth of the off-frequency band decreases with regard to the modulation depth of the envelope of the on-frequency band, CMR reduces and is determined by the coherence in the noise. There is a consistent decrease in CMR as modulation rate increases (Peters and Hall, 1994; Eddins and Wright, 1994). CMR using multiple modulation rates for the maskers depends upon the temporal information at both fluctuation rates; coherence at both rates leads to greater masking release than coherence at either modulation rate (Eddins and Wright, 1994).

For MDI, the reduction in sensitivity to SAM depends on the difference in modulation frequency for target and masker, with most masking occurring at identical modulation rates. The amount of MDI decreased as the common modulation frequency of the masker and signal increased (Yost and Sheft, 1989; Bacon and Konrad, 1993). In contrast, Bacon and Grantham (1989) reported a constant amount of modulation masking for common signal and masker frequencies

†Bacon and Konrad (1993) measured MDI for masker and target levels ranging, independently, from 40 to 80 dB SPL. When masker and target were equal in level, the amount of MDI increased somewhat with increasing level. When masker and target differed in level, the largest effects were reported when the target was close in frequency to the masker and the masker was more intense than the target.

of 4, 16, 64 Hz. The amount of MDI increases as the masker modulation depth increases (Yost and Sheft, 1989; Bacon and Konrad, 1993).

4.6.1 Are MDI and CMR related ?

The previous section discussed the properties of CMR and MDI independently. However, they are also related in at least two aspects:

1. **Within or across channel processing:** Both phenomena show a larger effect[‡] for a smaller carrier frequency separation, which implies a within channel effect. Both effects are also known to occur dichotically, which implies an across-channel effect. For these dichotic measurements, some reports mention a similar proximity effect as for the monaural measurements. This implies that both effects are mostly across channel effects, with a small contribution of peripheral interaction.

2. **Perceptual grouping;** CMR decreases as the difference in phase between the off-frequency and on-frequency band increases, whereas MDI remains constant. This difference in handling the phase suggests that MDI and CMR underlie different processes (Richards et al., 1997). In addition, CMR is very sensitive to a difference in the modulation rate of the envelope of the masker masking the target and the envelope of the off-frequency band. MDI is insensitive to differences in modulator phase and shows broad tuning for modulation frequencies. This is in contrast with perceptual grouping experiments. In contrast, adding cue tones reduces MDI. This indicates that MDI is at least to some extent attributable to difficulties in hearing out the pitch.

Both phenomena are subjected to across channel processing and perceptual grouping mechanisms. Since CMR results from 'popping out' and MDI results from fusion into one percept, it is likely that both phenomena are related. Originally, the differences in MDI and CMR were seen as quite large. However, the CMR and MDI concepts have been extended to other measures of temporal processing such as gap detection. In these cases, the addition of a modulated masker may lead to CMR or MDI. Adding a modulated masker will cause CMR when the threshold is high in the reference

[‡]Note that a larger effect indicates more masking release for CMR and a reduced sensitivity to SAM for MDI.

condition (only on-frequency band). when the signal produces a clear difference in the envelope. and when the signal and masker are gated asynchronously. When none of these conditions are fulfilled. MDI is more likely to occur Moore (1992).

4.7 Modelling temporal processing

Often, three different models are considered when describing temporal processing. These are the leaky integrator (Viemeister. 1979). the sliding temporal window (Moore et al., 1988). and the modulation filter bank (Dau et al., 1997b). The modulation filter bank differs from the other two models. since it includes a bank of overlapping bandpass filters acting on the envelope. whereas the leaky integrator and the sliding temporal window use a low-pass filter.

4.7.1 Leaky integrator

The leaky integrator. as described by Viemeister (1979). was primarily developed in order to describe the sensitivity to SAM for broadband carriers as a function of the modulation rate. This model consists of four successive stages: (1) 2000 Hz wide pre-detection filter. ranging between 4000 and 6000 Hz (first order Butterworth); (2) a half-wave rectifier; (3) low-pass filter (first order Butterworth) with a 3-dB cut-off frequency at 64 Hz; (4) a decision device generating a statistical representation of the incoming wave. The input of the decision device can be seen as the envelope of the original waveform. The decision device is often the parameter that distinguishes between studies. Strickland and Viemeister (1996) distinguished five different statistical representations of the fluctuating behavior of the signal:

1. The second central moment is given by the standard deviation of the output of the low-pass filter (Viemeister, 1979)

   $$\text{stdev} = \sqrt{\overline{x_t^2} - \overline{x_t}^2} \tag{4.4}$$

   in which x_t is the smoothed envelope given by the leaky integrator.

2. The coefficient of kurtosis corresponding to the fourth central moment, given by:

   $$\text{kurtosis} = \frac{\sum x_t^4}{(\sum x_t^2)^2}. \tag{4.5}$$

This describes the extent of the peak in a distribution (Hartmann and Pumplin, 1988). Values closer to 0 indicate a flatter, more uniform distribution.

3. The crest factor defined as the ratio between the maximum envelope and the RMS-power (Hartmann and Pumplin, 1988)

$$\text{crest} = \frac{max(x_t)}{RMS(x_t)} \tag{4.6}$$

4. The ratio between the maximum and the minimum amplitude of the envelope (Forrest and Green, 1987)

$$\text{maxmin} = \frac{max(x_t)}{min(x_t)} \tag{4.7}$$

5. The average magnitude of the slope of the envelope (Richards, 1992)

$$\text{slope} = \overline{|x_t - x_{t-1}|} \tag{4.8}$$

In the last couple of decades, the leaky integrator has proven to be a powerful tool in predicting psycho-acoustical performance in a range of experiments. Besides models for the TMTF for broadband noises, based on the standard deviation and max/min as decision device, the detection of a tone in narrow-band noise has been successfully modelled using the other three decision devices. Modulation masking experiments have also been modelled with reasonable success using the Crest factor as the decision device (Strickland and Viemeister, 1996). Trends such as the tuning for modulation rates, an increased masking with masker modulation depth, a change in threshold as a function of the signal phase relative to the masker phase, are best predicted using the max/min decision device. The Crest factor predicted some aspects of tuning better than the max/min device and also showed that the thresholds depended on the modulation depth of the masker was predicted. Berg (1996) showed that the leaky-integrator, using a single band as pre-detection filter, is useful for accounting for simplistic cases of CMR.

4.7.2 Sliding temporal window

The sensitivity to a signal is determined as a function of the duration between a preceding and following masker using a temporal notch instead of a spectral notch.

This is comparable to the determination of the auditory filter shape (see footnote section 4.3.3). Using an equation similar to Equation 4.2 determines the weighting of the window as a function of the gap. By sliding this window along the signal, the input becomes smoothed (Moore et al., 1988; Plack et al., 2002)[§]. In short, the model consists of four successive stages: (1) a dual-resonance nonlinear (DRNL) filter, which is a filter consisting of a linear and a non-linear pathway, filters the input into spectral bands (Meddis et al., 2001); (2) a non-llinearity; (3) a sliding temporal window with an equivalent rectangular duration[¶] of 10 ms; and (4) the decision device. The output of the model is generally based on the ratio of the signal plus masker and the masker alone. However, statistical devices, as mentioned in the previous section, may be more appropriate for experiments on the sensitivity to SAM.

The results of numerous psycho-acoustical experiments have been predicted correctly using the temporal sliding window. Evidently, data on forward masking and backward masking can be predicted accurately. Differences between increment and decrement detection are successfully modulated using a decision criterion based on the maximum slope of the temporal window output (Oxenham, 1997). The TMTF for low modulation rates can also be predicted correctly (Moore et al., 1989). However, subjects appeared to be less sensitive to high modulation rates than predicted by the model output. This can mainly be attributed to the relatively low cut-off frequency obtained by applying the temporal window, which can not adequately describe temporal processing for these high modulation rates.

4.7.3 Modulation filter bank

Dau et al. (1997b) introduced a model, referred to as the modulation filter bank, describing the temporal processing of sound. The model consists of four computational stages: (1) a filterbank, resembling the filtering occurring at the stage of the basilar membrane; (2) a half-wave rectifier followed by low-pass filtering at 1 kHz. The effects of adaptation, forward and backward masking, are simulated using five feedback loops with time constants ranging from 5 to 500 ms. Stationary

[§]A Matlab version of the sliding temporal window can be found on Dr. Plack's homepage: http://privatewww.essex.ac.uk/~cplack/temporal_window.html.

[¶]A detailed description of the temporal window shape is given by Plack and Oxenham (1998) and Oxenham and Plack (2000) input is transmitted logarithmically, whereas fast fluctuations are transmitted more linearly; (3) a modulation filterbank in which the resolution is limited by the addition of internal noise; (4) an optimal device to compare intervals to each other within the auditory task.

Results from numerous psycho-acoustical experiments have been correctly predicted, such as the TMTF, the TMTF as a function of the carrier bandwidth, and temporal integration. But also the modulation masking experiments as a function of modulation rate (tuning), masker modulation depth, and the critical bandwidth-concept for modulation rates as carried out by Houtgast (1989) were predicted correctly. Verhey et al. (1999) showed that a single channel analysis, based on one filter output, quantitatively describes most of the CMR-results for band widening paradigms (see section 4.4.1, CMR of the first kind), demonstrating that within-channel cues are strongly involved in this class of experiments.

4.8 Adaptation

Adaptation is the reduced sensitivity to a target due to pre-exposure to another signal. Some studies suggest that non-simultaneous masking can partly be attributed to adaptation (Duifhuis, 1973). However, adaptation is also known to occur for much longer durations, of which the temporary threshold shift (TTS) due to pre-exposure to loud signals is a well known example. From a physiological point of view, adaptation is reflected by the reduced discharge rate of the neurons lasting for periods up to 45 seconds (Javel, 1996). This adaptation is known to affect temporal integration (Zwislocki, 1960).

4.8.1 Adaptation to modulated signals

Psychophysical evidence

Psychophysical experiments show that pre-exposure to frequency modulated (FM) stimuli elevates thresholds for FM detection, but not for AM detection. Pre-exposure to AM signals elevates thresholds for AM detection but not for FM detection (Kay and Matthews, 1972; Tansley and Regan, 1979; Tansley and Suffield, 1983). These results suggest that modulation specific channels exist for FM and AM. This view is extended by suggesting that there are channels, specifically for upward and for downward sweeps in FM, since thresholds to an increment in FM are barely affected by downward sweeps in FM but are affected by upward sweeps in FM (Tansley and Regan, 1979; Gardner and Wilson, 1979). In addition, the threshold for the detection of an increment in intensity is affected by exposure to a tone of increasing intensity. A corresponding selective elevation occurs after adapting to a test tone of decreasing intensity (Tansley and Regan, 1979). Maximum adaptation occurs at modulation rates of about 16 Hz for AM and FM (Tansley and Suffield, 1983). The adaptation trajectory can be subdivided into two parts. The sensitivity to modulated signals drops almost immediately after exposure to modulated signals and remains constant after sufficient exposure (variable 8-30 minutes), indicating saturation. Following 10 minutes of adaptation, the threshold elevation recovered logarithmically within approximately 55 s to normal threshold values (Tansley and Regan, 1979). Bacon and Grantham (1992) measured modulation masking preceded and followed by a fringe, which contained the same contents as the masker. The differences for fringes that precede and follow the target, suggest that adaptation to SAM may occur at least partly for shorter durations of pre-exposure.

Physiological evidence

Physiological measurements show that the TMTF (Viemeister, 1979) can be determined for different locations on the auditory pathway. At higher levels of the auditory pathway, AM signals are presumably recoded in terms of synchronous neuromagnetic responses in the auditory cortex (*Mäkelä* et al., 1987; Kuwada et al., 2002). The cortex shows optimal activation for low modulation frequencies (<10 Hz), which drops for higher modulation rates, roughly comparable to the TMTF as demonstrated by Viemeister (1979). At lower levels of the auditory pathway, the amount of neural activity for modulated and unmodulated signals is similar, the actual AM is recoded in terms of the synchrony of neural activity (Møller, 1974; Schreiner and Urbas, 1986; Müller-Preuss et al., 1994). MTFs roughly show a low-pass function with a cut-off frequency that depends on the carrier frequency and neuron (Joris and Yin, 1992; Greenwood and Joris, 1996; Rees et al., 1986) or react bandpass-like tuned to the best modulation frequency close to 8-16 Hz (Müller-Preuss et al., 1994; Kuwada et al., 2002). In general, neurons do not respond solely to AM or FM. Roughly 30% respond exclusively to FM and 50% respond to both stimuli types (Rees and Moller, 1983; Eggermont, 2002).

Finally, neurons adapt to AM. Physiological experiments show that an AM-stimulus influenced the response to a second AM-stimulus. The same is found when both stimuli were FM-stimuli (Coombs and Fay, 1985; *Mäkelä* et al., 1987; Eggermont, 2002). However, when the type of stimulus differs, a preceding FM affects the response to AM more than a preceding AM influences the response to FM. Onset units (17 % of all neurons tested) which only fire at the beginning of each cycle have been reported that show a high degree of adaptation (Neuert et al., 2001; Eggermont, 2002).

Chapter 5

MDI for normal hearing and hearing-impaired subjects *

Abstract

This study evaluates whether modulation discrimination interference (MDI) differs for normal hearing and hearing-impaired subjects. The sensitivity to a change in the modulation depth of a 1.4 kHz sinusoid was measured. for six normal hearing and five hearing-impaired subjects, relative to reference depths of 0 (modulation detection), 0.18 or 0.30 (modulation discrimination) modulated by a sinusoid of 4, 8 or 16 Hz. The experiment was carried out while the target was flanked by either no maskers, non-modulated maskers, or modulated maskers. Maskers were situated at 500 Hz and 4 kHz and when modulated, modulated at the same modulation rate as the target. Modulated maskers reduced the sensitivity to a change in modulation depth. No clear differences were found between normal hearing and hearing-impaired subjects for the sensitivity to changes in modulation depth when the target was presented in the presence of modulated maskers. The sensitivity to a change in modulation depth was also reduced for normal hearing subjects by the presence of added non-modulated maskers. The reduced sensitivity to SAM due to the unmodulated maskers was not found for hearing-impaired subjects.

*This Chapter is part of a paper, which has been submitted to Ear & Hearing.

5.1 Motivation

The importance of the temporal structure of a signal has been discussed extensively in previous chapters. As described in Chapter 4, two aspects are important in the description of temporal processing in the presence of a modulated masker. Firstly, the sensitivity to a pure tone can increase when additional information can be derived from the envelope of the masker outside the auditory filter tuned to the target. An example is co-modulation masking release, referred to as CMR (Hall et al., 1984). Secondly, fluctuating maskers may decrease the sensitivity to SAM when modulated by a similar modulation rate. An example is modulation detection or discrimination interference, both referred to as MDI (Yost and Sheft, 1989).

In this study, MDI will be examined for normal hearing and hearing-impaired subjects. The literature only reports that modulation detection interference for hearing-impaired subjects is about the same as for normal hearing subjects (Bacon and Opie, 2002; Grose and Hall, 1994). A more complex task, such as modulation discrimination, could segregate the two groups and illustrate the problems hearing-impaired subjects experience in a fluctuating background noise.

5.2 Methods

5.2.1 Subjects

Six normal hearing and five hearing-impaired subjects participated in this study. The normal hearing subjects had thresholds smaller than 15 dB HL at octave frequencies from 0.125 to 8 kHz (*re.* ANSI, 1996). Stimuli were presented monaurally to a randomly assigned ear. Figure 5.1 gives the pure-tone thresholds for each of the individual hearing-impaired subjects. In general, the differences between a subject's two ears were small. Subjects with an asymmetrical hearing loss were tested using their better ear. The subject numbers were allocated according to the average hearing loss at the frequencies 1, 2, and 4 kHz. Lower numbers correspond to a smaller average hearing loss.

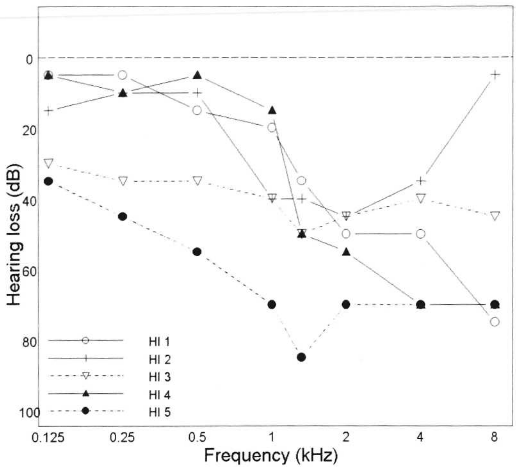


Figure 5.1: *Pure tone audiograms for the better ear of the five hearing-impaired subjects.*

5.2.2 Methods

Stimuli

The sensitivity to SAM was measured in three conditions. Firstly, the sensitivity to SAM for a target without maskers (probe alone; referred to as PA) was determined. Secondly the sensitivity to SAM for a target simultaneously presented with two non-modulated maskers (UNM) was examined. Thirdly the sensitivity to SAM for a target simultaneously presented with two modulated maskers (MOD) was tested. The task was to detect an increment in the modulation depth of the target. The target was a 1.4 kHz sinusoidal carrier, amplitude modulated by a 4, 8, or 16 Hz tone. The modulation depth of the reference signal was 0 (modulation detection), 0.18, or 0.30 (modulation discrimination). The maskers were presented at 0.5 kHz and 4 kHz. In the case of modulated maskers, the maskers were modulated using the modulation rate of the target. The modulation depths of the maskers were, if modulated, similar to that of the target, but were roved by 2 dB (in modulation depth) to avoid possible effects of perceptual grouping based on a common modulation depth. The phases of the envelopes of target and masker were unrelated, starting at a random phase. The stimuli were 1000 ms in duration, including cosine squared rise/fall times of 15 ms. The inter-stimulus intervals were equal to 200 ms. All carriers were presented at the most comfortable level (MCL) according to the relationship between MCL and hearing threshold as described by Lyregaard (1988) (see Table 5.1 for the individual MCL-levels).

	Level of carrier (dB SPL)		
subject	0.5 kHz	1.4 kHz	4 kHz
NH	57	53	46
HI-1	65	71	72
HI-2	62	74	64
HI-3	75	79	67
HI-4	59	79	87
HI-5	85	101	87

Table 5.1: *The presentation level (in dB SPL) for each subject for each carrier frequency.*

Procedure

Subjects were seated in a sound-attenuating booth. The sensitivity to SAM was determined using a 3I-3AFC set-up, with a 1-up 3-down rule, converging to the 79.1 % point of correct responses on the psychometric curve (Levitt, 1971). Correct-answer feedback was given after each trial by a light on the computer screen. Initially, the step size for a change in the modulation depth (Δm) was 3 dB in units of $20log_{10}(\Delta\text{m})$. After two reversals, the step size was reduced to 1 dB. The threshold was determined by taking the average of the last 8, of 12, reversals. Classical MDI (MDI_{unm}) was determined using :

$$MDI_{unm} = 20\log_{10}(\Delta m_{mod}) - 20\log_{10}(\Delta m_{unm}). \tag{5.1}$$

An alternative method to determine the effect of modulated maskers is given by :

$$MDI_{pa} = 20\log_{10}(\Delta m_{mod}) - 20\log_{10}(\Delta m_{pa}). \tag{5.2}$$

Here the subscripts *pa*, *unm* and *mod* denote the sensitivity to a change in SAM without maskers, in the presence of unmodulated maskers, and in the presence of modulated maskers, respectively.

After 2 hours of training, data-collection started. Measurements were carried out in test and in retest. When test and retest deviated by more than 3 dB a third test was conducted. The two closest estimates were used for further statistical analysis. The modulation depth was restricted to 1 (0 dB) to prevent over-modulation.

Apparatus

Stimuli were generated digitally at a sampling rate of 25 kHz using a Tucker-Davis Technology (TDT-II) system. The stimuli passed a 16-bit DA-converter (DA3-2) and an anti-aliasing filter (FT6; cut-off frequency 8 kHz). Presentation levels were controlled by a programmable attenuator (PA4) and summed by a summer (SM3). These signals passed a headphone buffer (HB6) before presenting it to headphones (Telephonics TDH 39-P).

5.2.3 Statistical analysis

Examining the accuracy of measurement

The accuracy of measurements, like those described in this Chapter, can be examined using Cronbach's alpha (Cronbach, 1951).

$$\alpha = \frac{k}{k-1}[1 - \frac{\sum_{i=1}^{k} \sigma_{x_i}^2}{\sigma_x^2}] \tag{5.3}$$

where k indicates how often tests are carried out, $\sigma_{x_i}^2$ denotes the variance of the score on test i and σ_x^2 the variance of the sum of the scores on all tests. If the performance of an individual is to be determined using a single condition. Cronbach's alpha should usually be at least 0.85 (Frisbie, 1988). However, if groups of individuals are to be compared, a lower level is acceptable. Values of Cronbach's alpha of 0.7 and above were regarded as good. Cronbach's alpha was calculated using SPSS10.0.

Correction for attenuation

The maximum proportion of variance in a variable, y, which can be explained by another variable, x, depends on the accuracy with which x and y can be measured.

The proportion of variance, r_c^2, that can be explained by a model, if measurements were obtained without measurement errors can be calculated using the correction for attenuation given in Eq. 5.4.

$$r_c^2 = \frac{r^2}{\alpha_x \alpha_y} \tag{5.4}$$

where r is the observed correlation between the values of x and y and α_x and α_y are the values of Cronbach's α for the measurements of x and y, respectively.

Linear Mixed Effects model

In order to examine the effects of a parameter on, for instance, the sensitivity to a change in SAM, a linear mixed effects model with the sensitivity to a change in SAM as the explanatory variable and the parameters as the dependent variables was fitted, using S-PLUS. A random intercept and slope were included for each subject (Pinheiro and Bates, 2000). Hierarchical models can be compared using the likelihood ratio test, which is comparable to the F-test used in linear regression. The level of significance was determined using a linear mixed effects model, unless stated otherwise.

5.3 Results

There was no significant learning effect (paired t-test: p-value $= 0.38$) after two hours of training. The accuracy of the measurements was quite high according to Cronbachs alpha ($\alpha = 0.97$).

The main effects and interaction terms were studied using linear mixed effects models. Since subjects differed significantly (Linear Mixed Effects model (LME), p-value<0.0001) individual data averaged over modulation rate are presented for normal hearing subjects in Figure 5.3 and for hearing-impaired subjects in Figure 5.3. Error bars represent the standard deviation over test sequence and modulation rate. Both panels show that, for normal hearing and hearing-impaired subjects, the sensitivity to a change in SAM is highest when measured without flanking tones (open circles), decreases when pure tone flankers are added (asterisks) and reduces seriously by the addition of modulated maskers (filled triangles). Modulation of the

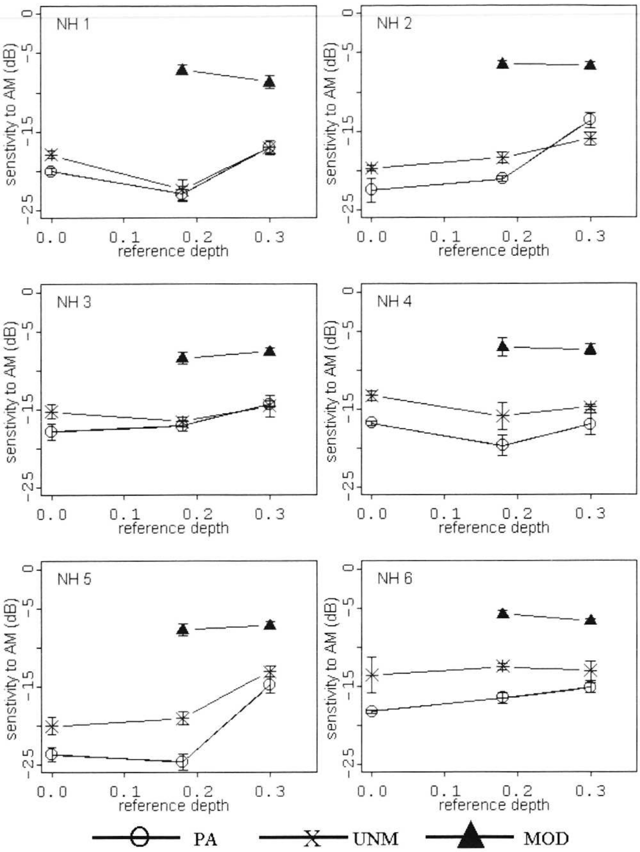


Figure 5.2: *Sensitivity to a change in SAM for normal hearing subjects as a function of the reference depth. Circles give the results for experiments without maskers (PA), asterisks for the experiments with non-modulated maskers (UNM) and triangles for the experiments with modulated maskers (MOD).*

maskers results in a reduction of the sensitivity to a change in SAM of approximately 7 to 10 dB (ranging from 3 dB to 15 dB).

5.3.1 Normal hearing subjects

The sensitivity to a change in SAM for normal hearing subjects (Figure 5.3) was barely affected by modulation rate, since neither main effect (LME, p-value $= 0.9$) nor interaction effects with reference depth and masker type (LME, p-value > 0.1) were significant. Therefore, data were averaged across the three modulation rates.

There was a significant main effect of reference depth (LME, p-value < 0.0001). The sensitivity to a change in SAM was not altered by increasing the reference depth from 0 to 0.18 (paired t-test: p-value $= 0.39$). However, increasing the modulation depth from 0.18 to 0.30 resulted in a reduced sensitivity to SAM (paired t-test: p-value < 0.0001). The main effect of masker type was also significant (LME, p-value < 0.0001). Adding non-modulated maskers significantly reduces the sensitivity to a change in SAM for reference depth 0 and 0.18 (paired t-tests, p-value < 0.001), whereas adding unmodulated maskers does not affect the sensitivity to a change in SAM for reference depth 0.30 (paired t-tests, p-value $= 0.46$). Adding modulated maskers rather than non-modulated maskers significantly reduces the sensitivity to a change in SAM at reference depths 0.18 and 0.30 (paired t-tests, p-value < 0.001). Hence, the interaction effect of reference depth and masker type was also significant (LME, p-value $= 0.0023$).

5.3.2 Hearing-impaired subjects

The results for hearing-impaired subjects are given in Figure 5.3. There are no significant main or interaction effects of the modulation rate. Therefore, the data were averaged over the three modulation rates. The sensitivity to a change in SAM decreases significantly as reference depth increases (LME, p-value < 0.0001). Masker type was also a significant main effect (LME, p-value < 0.0001). Adding non-modulated maskers to the target did not significantly reduce the sensitivity to a change in SAM. However, modulating these maskers significantly reduced the sensitivity to a change in SAM (paired t-tests, p-value < 0.001). This reflects the significant interaction effect of reference depth and the masker type (LME, p-value < 0.015).

5.3.3 Comparison between normal hearing and hearing-impaired subjects

In a separate analysis the data from the normal hearing and hearing-impaired subjects were pooled with hearing-impairment as an additional factor. Hearing-impairment was not a significant main effect. However, the interaction between impairment and masker type was significant (LME, p-value $= 0.001$), due to the

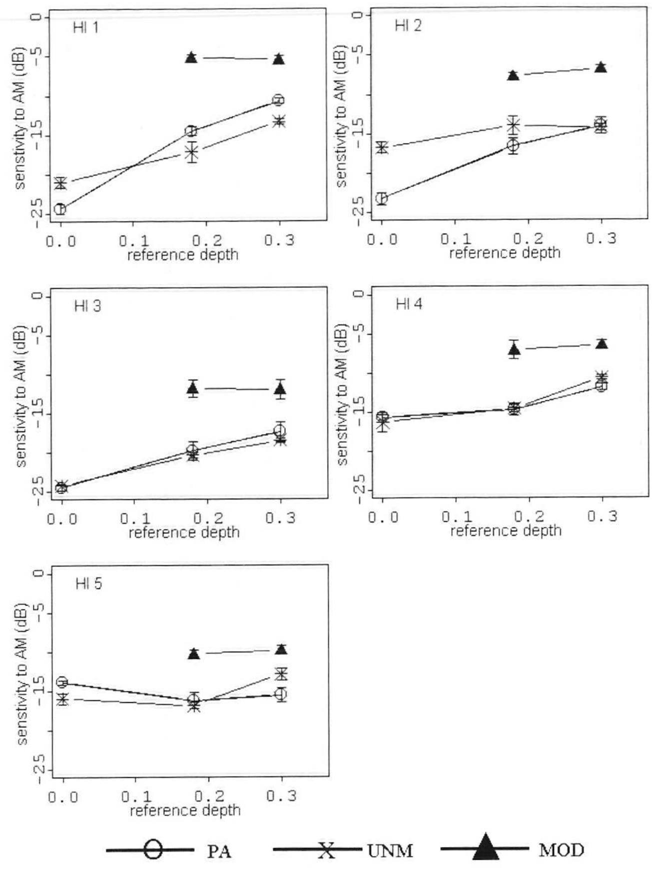


Figure 5.3: *Similar to Figure , but now for hearing-impaired subjects.*

reduced sensitivity to a change in SAM as a consequence of non-modulated maskers. This was found to be significant for normal hearing subjects (paired t-tests, p-value < 0.001) but not for hearing-impaired subjects (paired t-tests, p-value $= 0.27$).

5.4 Discussion

In this study, the average sensitivity to SAM for normal hearing subjects equals -20 dB, compared to -23 dB reported in literature at similar exposure levels (Kohlrausch et al., 2000). However, this sensitivity is known to differ among subjects. In addition, the sensitivity to SAM, using sinusoidal carriers, increases with sensation level (Kohlrausch et al., 2000).

Increasing the reference depth reduces the sensitivity to SAM. In this study, the Weber fraction for normal hearing subjects is similar for reference depth 0.18 and 0.30. However, compared to von Fleischer (1980), the Weber fraction from this study is larger (0.85 vs. 1.59). Except for the exposure level, no clear differences in experimental procedure could account for these differences. In addition, the reduced sensitivity to SAM as a consequence of increasing reference depth was similar to that reported by Wakefield and Viemeister (1990) for wideband noises and by Yost and Sheft (1994) for pure tone carriers.

The absence of a main effect relating to hearing capacity has also been reported in previous studies. Hearing-impaired subjects with relatively flat hearing losses are more sensitive than, or as sensitive as, normal hearing subjects (Bacon and Gleitman, 1992). Using a broadband carrier indicated that a reduced sensitivity to SAM can partly be attributed to audibility, since reducing the audibility by filtering for normal hearing subjects results in performance which is more in line with the performance of hearing-impaired subjects (Bacon and Viemeister, 1985).

In line with the results reported in this manuscript, the sensitivity to SAM is reduced by adding maskers for modulation detection (Yost and Sheft, 1989) and modulation discrimination (Moore et al., 1991; Moore and Shailer, 1992). The sensitivity to a change in SAM reported in this manuscript is somewhat lower than in previous studies. Using modulated maskers reduced the sensitivity to SAM up to 15 dB (on average 8 dB), which is similar to the 8 dB reported by Wakefield and Viemeister (1990); Moore et al. (1991); Moore and Shailer (1992). Most studies differ in the number of maskers added or in the reference depth of the target or masker. Yost and Sheft (1994) showed that the amount of MDI decreases as reference depth increases. This is mainly due to a larger loss of sensitivity to a change in SAM when the target is presented in the presence of non-modulated maskers. Results from this study reported similar findings. Bacon and Moore (1993) measured MDI using one masker, modulated with a modulation depth of 0.5. This resulted in a smaller interference (6 dB). The amount of interference for normal hearing and hearing-impaired subjects was similar (Grose and Hall, 1994). These data are in line with the data reported in this manuscript. Yost and Sheft (1989) reported a higher interference: 10 dB for masker carrier frequencies higher in frequency and 7 dB for maskers lower in frequency. However these data examined modulation detection and were obtained with a single masker, which leads to lower MDI-values. Similar results have been reported for the largest frequency separation (Mendoza et al., 1995b).

For normal hearing subjects, adding non-modulated maskers reduces the sensitivity to a change in SAM with regard to the condition without maskers. These effects have also been reported in other studies (Yost and Sheft, 1989; Bacon and Moore, 1993) and are, at least partly, attributed to the energy from the maskers falling in the auditory filter of the target. The sensitivity of hearing-impaired subjects to a change in SAM is not reduced when non-modulated maskers are added. This is surprising, since a larger interference is to be expected, based on the broader auditory filters for hearing-impaired listeners and the presentation at well audible levels. This implies that either the reduction in sensitivity to SAM caused by adding pure-tone maskers does not result from energy falling in the target filter, or a within channel effect is absent for hearing-impaired subjects.

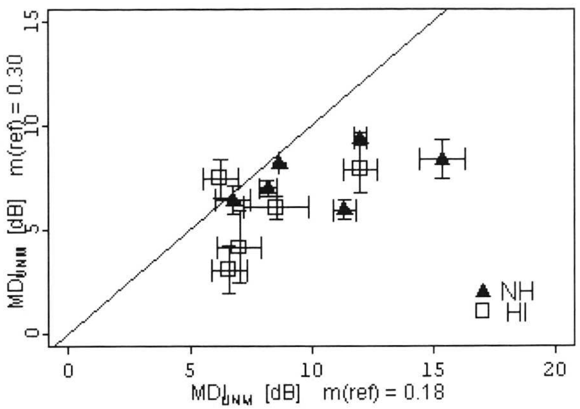
The reduced sensitivity to a change in SAM resulting from additional non-modulated carriers is larger for reference depth 0 than for 0.18 and is absent for 0.30. Given that the level of the target and maskers remained constant[†], these data provide additional evidence that pure tone maskers interfere with the modulation discrimination process at a more central level than the peripheral processing. This cannot be attributed to the energy added by the non-modulated flankers. Based on the modulation filter bank model (Dau et al., 1997b), this interference may be incorporated at two levels. Firstly, the internal noise, limiting temporal resolution, may depend on the total RMS-level of the signal. Secondly, at the level of the modulation filter bank, the energy of the DC-component (0 Hz) partly falls in the modulation filter tuned to the modulation rate of the stimulus. Although the main effect is not significant, a weak effect of more interference is found for the non-modulated masker for a 4 Hz modulated target than for a 8 Hz target (paired t-test; p-value $= 0.014$; 1.2 dB). However, the interference due to non-modulated maskers is not significantly different between the modulation rates 8 and 16 Hz (paired t-test; p-value $= 0.36$). Alternatively, these effects may be explained in terms of

[†]Modulated signals were not scaled to equal RMS or using the correction formula suggested by Houtgast (1989). However, it is unlikely that this would have influenced the results to a large extent. A more extensive overview on the impact of intensity scaling is given by Mendoza et al. (1995b). In addition, it was shown, using a matching paradigm, that modulated signals are not perceived as louder than non-modulated signals (Moore et al., 1999).

the multiple looks model. Since the overall duration is constant, lower modulation rates result in fewer cycles during which a change in SAM could be detected and thereby reducing performance. However, in this case, either a main effect of the modulation frequency would be expected. A higher sensitivity to SAM for 8 Hz than for 4 Hz, but a similar sensitivity to SAM for 8 and 16 Hz would have been expected. Moreover, the critical duration for modulation discrimination without flankers corresponds to 4-5 cycles (Lee and Bacon, 1997). Hence, the stimulus was just long enough to fit this critical duration, for a modulation rate of 4 Hz. For 8 and 16 Hz, the duration of the signal clearly extends the critical duration. In addition, the critical duration was assessed for measurements in the PA task measuring modulation discrimination without simultaneously presented maskers. For more complex tasks, such as competing backgrounds given by modulated or unmodulated maskers, the critical duration may increase.

The reduction in sensitivity to a change in SAM due to modulated maskers instead of non-modulated maskers is referred to as MDI_{UNM}. The reduction in sensitivity to a change in SAM for modulated maskers relative to the situation without maskers is referred to as MDI_{PA}. The top panel in Figure 5.4 shows a scatter plot between MDI_{UNM} with reference depths of 0.18 and 0.30. The bottom panel in Figure 5.4 shows a similar scatter plot for MDI_{PA} using reference depths 0.18 and 0.30. The MDI-values for both figures have been averaged across the three modulation rates. The diagonal line indicates the points for which the amount of MDI would be equal for reference depths 0.18 and 0.30. Most data-points are on the right-hand side of this line, indicating that MDI at m=0.30 is usually smaller than at m=0.18. For both reference depths, the amount of MDI_{UNM} is not significantly different between normal hearing and hearing-impaired subjects (t-test: p-value = 0.11). However, MDI_{PA} is significantly smaller for hearing-impaired than for normal hearing subjects, independent of the reference depth used (t-test: p-value < 0.01 [$m_{ref} = 0.18$], p-value $= 0.02$ [$m_{ref} = 0.30$]). The two groups are clearly segregated.

In order to identify the modulated target, its modulation depth needs to be fairly large. To avoid overmodulation, the maximum modulation depth of an amplitude modulated signal is 1. The detection-space of SAM is considered linear on a logarithmic scale (dB-transformation) and subjects could always identify the correct interval of the fully modulated signal. However, perceptually this may not



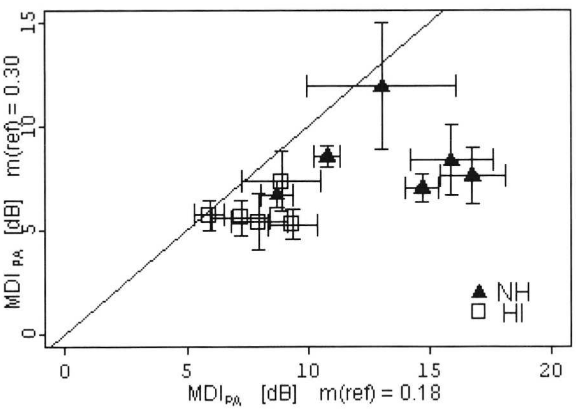


Figure 5.4: *Scatter plots of the interference obtained using two different reference depths for two different methods of determining the interference caused by modulated maskers. Each symbol (squares normal hearing; triangles hearing-impaired subjects) gives the amount of interference found at a reference depth of 0.18 (abscissa) versus the interference found for a reference depth of 0.30 (ordinate). The top panel gives the classical MDI-measure (MDI_{UNM}). The bottom panel presents an alternative measure with regard to the condition without maskers (MDI_{PA}). Error bars show the standard deviations over modulation rates for test and retest. The sloping line shows the points at which the interference at reference depth 0.18 would be equal to 0.30.*

necessarily be true. Imagine a detection space that is build up in steps of just noticeable differences and that this detection space is not linear over the whole range of modulation perception (*i.e.* in equal dB steps), but, for instance compressive.

The steps in this detection space reduce and discrimination of SAM would occur for smaller differences (in dB) that can probably not be segregated when using too large step-sizes in the up-down procedure. Thresholds would be similar across subjects and MDI would be primarily based on the sensitivity of the subject to SAM with non-modulated maskers. Evidence supporting this kind of non-linearity in modulation perception has been reported in literature (von Fleischer, 1980; Ozimek and Sek, 1988; Wakefield and Viemeister, 1990). Ozimek and Sek (1988) measured the sensitivity to a change in SAM for pure tone carriers as a function of increasing reference depth. The sensitivity to a change in SAM decreases with increasing reference depth, according to the Weber law. However, for reference depths above 75%, the sensitivity to a change in SAM increased, indicating a non-linearity. If the reference modulation depth increases from 0.18 to 0.30, the change in sensitivity also increases for the discrimination task without flankers and with non-modulated maskers. The sensitivity to a change in SAM using modulated maskers remains unaltered (difference in thresholds is not significantly different from 0; t-test p-value $= 0.9$). This indicates that the differences could have been restricted by the limited detection space. The largest modulation depth required equals -2.7 dB (m=0.73; HI 1) for a reference depth of 0.18; the average equals -4.6 dB (m=0.59). For reference depth 0.30 these values are -1.6 dB (m=0.83; HI 1)[‡] and -3.0 dB (m=0.71). Well above the range for which the Weber fraction was found to be constant as a function of increasing reference depth. These depths are relatively high and some subjects require a fully modulated signal in order to detect the differences in modulation depth. So, lack of differences in MDI-values may be induced by the high thresholds and does not necessarily mean that the thresholds would not differ if somehow thresholds were not limited to $m = 1$.

[‡]An average of -1.6 dB may seem impossible to the reader since signals were not overmodulated and step sizes of 1 dB were taken. However, the step sizes were taken from the differences in modulation depth, which was added to the reference depth. For example, a modulation depth of -1.6 dB equals 0.83, the difference in modulation depth from the reference equals 0.53 or -5.5 dB. Adding 2 dB (step size = 1 dB) results in -3.5 dB or 0.67, which together with the reference depth (0.3) results in 0.97, which is just smaller than the maximum modulation depth.

5.5 Conclusions

The sensitivity to a change in the modulation depth of a SAM-pure tone has been determined with regard to reference depths of 0 (modulation detection), 0.18 and 0.30 (modulation discrimination). The experiments are carried out for a target without maskers (PA), in the presence of non-modulated maskers (UNM), and in the presence of modulated maskers (MOD). The results can be summarized as follows:

1. the amount of modulation discrimination interference (MDI), is comparable for normal hearing and hearing-impaired subjects. Differences in MDI can be attributed to differences in the UNM-task.

2. the UNM task shows a reduced sensitivity to SAM for normal hearing subjects at reference depths 0 and 0.18, but not for 0.30, relative to the PA task. Thresholds for hearing-impaired subjects remain unaltered when non-modulated maskers are added.

3. the amount of MDI is mainly determined by the sensitivity to UNM.

Chapter 6

Amplitude modulation matching for subjects with normal and impaired hearing

Abstract

This study focused on supra-threshold modulation perception. Matching experiments were carried out for normal hearing and hearing-impaired subjects to study the effects of altering different parameters (bandwidth, center frequency, or sensation level) on the perception of amplitude modulation. Signals were modulated by a sinusoidal amplitude modulation of 8 Hz using reference depths of either m=0.5 or m=0.7. Carriers were presented at 10 or 25 dB SL, using center frequencies of 1 kHz or 4 kHz. The bandwidth was either narrower or wider than the critical bandwidth. The bandwidths of the narrowband signal at 4 kHz and the wideband signal at 1 kHz were chosen so that the absolute bandwidth was equally large in Hz. The results could be described reasonably well in terms of the differences in the amount of slow inherent fluctuations in the target and reference signals. The differences in the growth of loudness do not correlate strongly with the differences in

modulation depth for target and reference. Hearing-impaired subjects adjusted the modulation depth of a target less accurately to the modulation depth of a reference. There was an overall trend for hearing-impaired subjects to adjust the modulation depth of the target to be higher than the modulation depth of the reference. In the Appendix, the results of normal hearing subjects were modelled using three different models known to describe temporal processing in the human ear.

6.1 Motivation

The primary motivation for this study was largely empirical as supra-threshold modulation perception and the underlying influence of parameters such as the bandwidth, center frequency, and sensation level have, to the knowledge of the author, not been previously investigated in a systematic way. In addition, a previous study measuring modulation discrimination interference (MDI) suggested that modulation perception for highly modulated stimuli differs from modulation perception for lower modulation depths (Chapter 5). These findings may be important to speech intelligibility, since speech contains these kinds of strong modulations.

6.2 Methods

6.2.1 Participants

In total, six normal hearing (average age 26 years [23-32]) and five hearing-impaired subjects (average age 18 years [17-20]) participated in the experiments. In the case of an asymmetrical hearing loss, the hearing-impaired subjects were tested in their better ears. All normal hearing subjects had pure-tone thresholds smaller than 15 dB HL (*re.*: ANSI, 1996) at octave frequencies from 0.125 kHz to 8 kHz. Hearing-impaired subjects were ranked according to the average loss at the two frequencies of interest (1 kHz and 4 kHz). The pure-tone audiograms of the test ear of the hearing-impaired subjects are given in Figure 6.1. The hearing-impaired subjects were volunteers and were paid for their contribution. They were recruited from a school for the hearing-impaired. Normal hearing subjects were coworkers from our department.

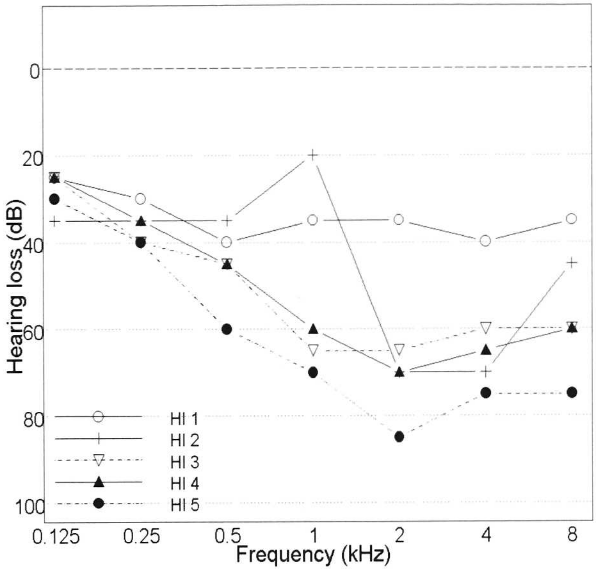


Figure 6.1: *Pure tone thresholds of the tested ear for the hearing-impaired subjects.*

6.2.2 Study design

Subjects were seated in a sound attenuating booth, facing a computer screen. They were asked to adjust the modulation depth of the target until it matched the modulation depth ('wobbling sensation') of a certain reference signal using a method of adjustment.

To initiate a trial, the subject clicked the mouse button. The first interval always contained the reference stimulus and the second interval the target. Reference and target were marked by a light signal. The modulation depth of the target could be changed by moving the scroll-wheel on a mouse and started with a randomly chosen modulation depth between the reference depth and 0 dB. The overall range, in which the modulation depth could be varied, was 30 dB (minimum depth equaled -30 dB; m=0.03). Each step of the scroll-wheel changed the modulation depth by 0.3 dB. After at least four experimental trials, the subject was allowed to press an OK-button, indicating that target and reference were perceived as equally modulated. This match was generally reached after 10 pairs of stimuli (lasting approximately 1 minute). Each test-session lasted about 10 minutes.

The difference between the modulation depth of the matched target and reference is referred to as the difference in physical modulation depth for an equal modulation perception (dm-EMP: difference in modulation depth for an Equal Modulation Perception) and is calculated by taking the differences in decibels between the matched modulation depth (m) and the reference modulation depth (m_{ref}) as given by Eq. 6.1:

$$\text{dm-EMP} = 20\log_{10}(m) - 20\log_{10}(m_{ref}). \tag{6.1}$$

Hence. the dm-EMP gives the number of decibels the matched target was adjusted above (dm-EMP > 0) or below (dm-EMP < 0) the reference depth.

6.2.3 Stimuli

The carrier of the target differed from the reference by either increasing or decreasing the bandwidth, center frequency, or sensation level. Table 6.1 describes the collection of data for the three parameters. Each parameters leads to 16 conditions for the individual listener, resulting in 48 conditions (3×16). All conditions were repeated four times, resulting in 192 data-points (4×48). The first presentation of a condition was regarded as a training and excluded from further analysis (remaining 144).

	change up/down	m_{ref} 0.5/0.7	f_c 1 kHz/4 kHz	SL 10/25	BW narrow/wide	Number of conditions
Bandwidth	2	2	2^+	2^+		16
Center frequency	2	2		2^+	2^+	16
Sensation level	2	2	2^+		2^+	16

Table 6.1: *Conditions for each parameter for each subject. Cumulative distributions have been taken for the conditions denoted by a plus.*

Signals were modulated using a sinusoidal amplitude modulation rate of 8 Hz with a reference depth of 0.5 or 0.7, starting at a random phase. Carrier signals were noise-bands with center frequencies of 1 kHz or 4 kHz. There were two noise bandwidth conditions at each frequency, one smaller than, and one larger than the estimated critical bandwidth for normal hearing subjects at that frequency (Zwicker, 1961). At 1 kHz, the bandwidths were 1/6 octave [944 Hz-1059 Hz] and 1 octave

[707 Hz-1414 Hz]. At 4 kHz, the bandwidths were 1/4 octave [3668 Hz-4362 Hz] and 1 octave [2828 Hz-5657 Hz]. The narrow-band noise at 4 kHz was chosen to be approximately equal in absolute bandwidth (*i.e.*, Hz) to the wideband noise at 1 kHz. Signals were presented at 10 or 25 dB SL.

The stimulus duration was 625 ms, 15 ms raise/fall times and 595 ms steady state, in order to fulfil the requirements of the critical duration of 4 to 5 periods of the modulation frequency (Lee and Bacon, 1997). The inter-stimulus interval equalled 200 ms. Noise bands were generated by adding sinusoids with constant amplitude and random phase, spaced by 0.001 times the center frequency. These noise bands were digitized at a sampling rate of 16 kHz. Signals were scaled to equal RMS-level and played back by the DA-converter of a TDT-II system, passed through an anti-aliasing filter ($f_{cut-off}$ = 8 kHz) and presented monaurally by SONY MDR-V900 headphones (using an interacoustics AC5 audiometer as headphone amplifier).

6.2.4 Simulating loudness recruitment

In order to simulate the effect of loudness recruitment for normal hearing subjects, matching experiments were carried out with signals presented in a background of pink noise. The pink noise was used to simulate reduced audibility resulting from hearing loss. Loudness functions for normal hearing subjects become similar to those of hearing-impaired subjects, in the sense that signals presented near the detection threshold (in noise) were almost inaudible, whereas almost normal loudness was achieved when the signals were presented 20 dB above the detection threshold in noise (Schlauch et al., 1998). Detection thresholds for the pink noise and signals presented in pink noise were determined using a 2I-2AFC procedure, in which the sensation level was decreased after two consecutive correct responses and increased after 1 incorrect response, tracking the level of 70.7% correct responses (Levitt, 1971). Thresholds were elevated by approximately 20 dB, offering a slightly lower exposure level than the average exposure level (in dB SPL) for the hearing-impaired subjects in this study. Matching experiments in pink noise were carried out for a change in sensation level, while keeping the bandwidth and center frequency constant.

6.2.5 Presentation of the results

Presenting data using cumulative distributions, rather than summary statistics, gives more insight into the structure of the data, particularly when comparing two or more groups. The data from the present experiments were combined into cumulative distributions, as indicated by Figure 6.2, to determine four parameters describing modulation perception:

1. effect: experiments were conducted for increases and decreases in bandwidths, center frequencies, or sensation levels. Increasing or decreasing one of these parameters may result in two different cumulative distributions. The difference between these two distributions, as given by the medians, will be known as the effect-size. The larger the effect-size, the larger the effect on modulation perception induced by the change of a signal parameter.

2. shift: the shift determined the percentage of responses adjusted to be higher than the reference depth. Two different cumulative distributions are given in the upper left corner of Figure 6.2. The percentage of targets adjusted higher than the reference depth is given by intersection with the ordinate. Normally, the shift is expected to equal 50%, since an increased signal parameter is expected to result in a similar result, but opposite effect as for a decreased signal parameter. Percentages larger than 50% indicate that subjects tended to adjust the modulation depth to be larger than the reference depth.

3. slope: the slope of the cumulative distributions is determined by dividing 100 by the difference in maximum and minimum dm-EMP. Figure 6.2 (bottom-left) indicates three cumulative distributions. The slope is given by straight lines connecting the maximum and minimum modulation depth. The dashed lines indicate similar slopes, whereas the dotted line indicates a more shallow slope. The latter indicates that subjects were less accurate for that condition. If the matches equal 1, as indicated by the small cumulative distribution, the slope was determined by dividing the percentage at which this occurred by the difference in modulation depths. The larger the slope, the more accurately subjects could match the modulation depth.

4. ceiling: the ceiling effect was determined by the percentage of presentation, in which the target was adjusted to be as fully modulated. Figure 6.2 (bottom-right) indicates two conditions. One curve is given along the full ordinate, the other up to 55% after which it reaches its maximum modulation depth. This ceiling effect may affect the slope and effect size since it may indicate that subjects tried to increase the modulation depth, while this was limited by signal limitations (m=1).

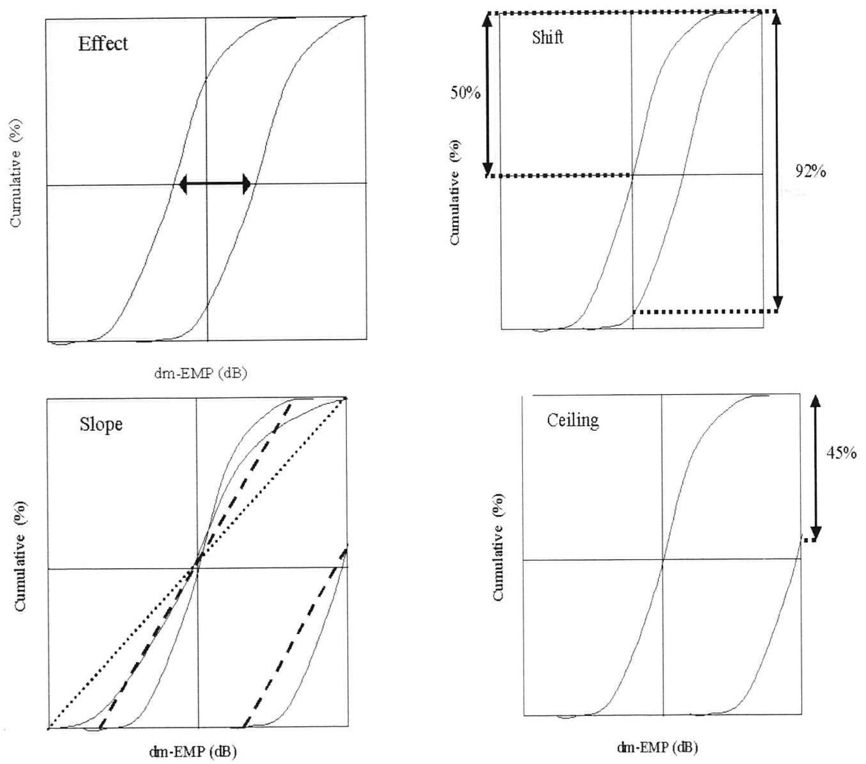


Figure 6.2: *The four parameters used to describe modulation perception. Effect is given by the distance between the cumulative distributions along the abscissa. Shift is given by the percentage adjusted larger than the reference depth (dm-EMP > 0). Slope is given by the averaged slope of the cumulative distributions. Ceiling is the percentage of matched signals that could not be increased further in modulation depth.*

6.3 Results

6.3.1 Normal hearing subjects

The purpose of the matching experiment is to investigate how changing a parameter influences modulation perception. The results for normal hearing subjects are given by cumulative distributions in Figure 6.3 for changing the bandwidth (BW), the center frequency (F_c), or the sensation level (SL). Additional information on these parameters for normal hearing subjects is given in Table 6.2.

For the signals to be perceived as equally modulated, the modulation depth of a wideband target (triangles Figure 6.3) had to be adjusted lower than the modulation depth of a narrowband reference (Figure 6.3: 80% of the observations indicate dm-EMP < 0). The modulation depth of a narrowband target was generally adjusted to be higher than the modulation depth of a wideband reference (75% of the observations, given by the squares, indicate dm-EMP > 0). In other words, the modulation depth of a wideband target needs to be decreased to achieve the same sensation of the modulation depth as the narrowband reference. The distance between the triangles and squares, measured along the median, is similar (paired t-test,p-value $= 0.15$) for the reference depths 0.5 (open symbols) and 0.7 (solid symbols), indicating that the size of the effect is not strongly dependent on reference depth. The *shift* is not significantly different from 50% (t-test; p-value $= 0.40$). Furthermore, the slope and ceiling indicate accurate matchings, suggesting that the results are reliable. This is supported by Cronbach's α, which equalled 0.78.

The matching experiments in which the *center frequency* was altered, given in the upper right panel of Figure 6.3, indicate that the modulation depth of a noise with center frequency 4 kHz (target) was adjusted to be lower in modulation depth than the modulation depth of a 1 kHz reference noise-stimulus (median of triangles indicate dm-EMP < 0). The modulation depth of a 1 kHz target noise signal was adjusted to be larger than the modulation depth of a 4 kHz reference (medians of the squares indicate dm-EMP > 0). The effect size, is smaller (paired t-test, p-value $= 0.045$) for reference depth 0.7 (solid symbols) than for 0.5 (open symbols). Other parameters given in Table 6.2 indicate, by showing large slopes and low ceiling effects, that matches were quite accurate. This is partly supported by Cronbach's α, which is very large ($\alpha = 0.88$).

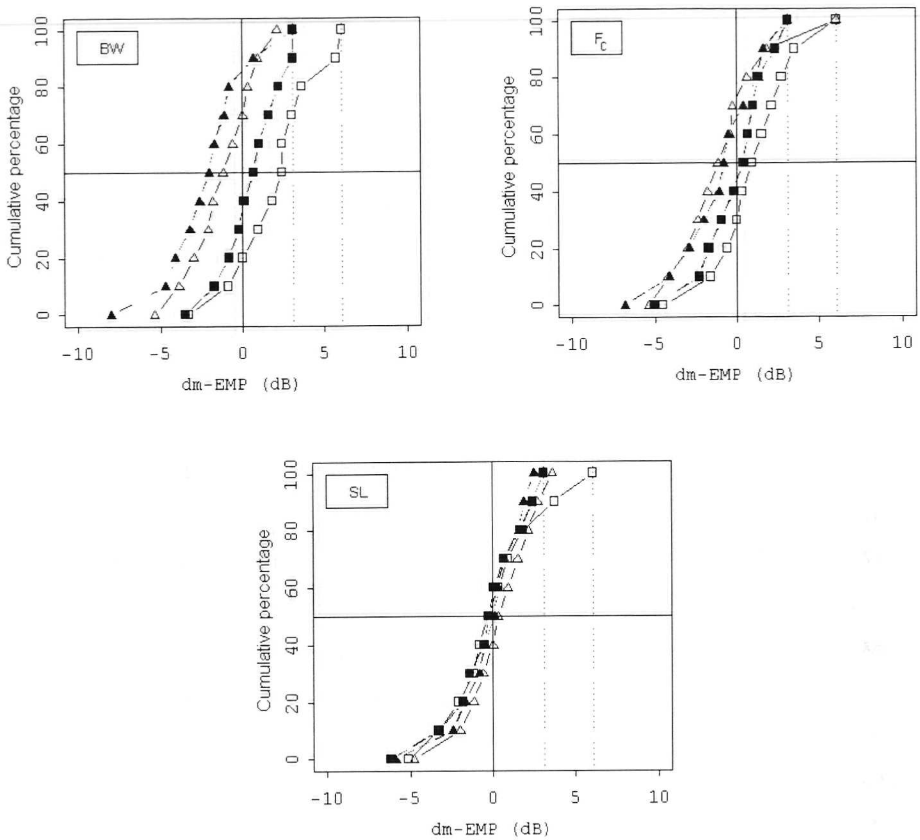


Figure 6.3: *Cumulative distribution for the data from normal hearing subjects. Experiments carried out with a reference depth of 0.5 are represented by open symbols; solid symbols represent data measured using a reference depth of 0.7. The cumulative curves represent :*

BW: triangles → reference narrowband, target wideband.
squares → reference wideband, target narrowband.
F_c: triangles → reference 1 kHz, target 4 kHz.
squares → reference 4 kHz, target 1 kHz.
SL: triangles → reference 10 dB SL, target 25 dB SL.
squares → reference 25 dB SL, target 10 dB SL.

The lowest panel of Figure 6.3 (SL) displays the results of the matching experiments in which the *sensation level* increases from 10 dB SL to 25 dB SL (triangles) or decreases from 25 dB SL to 10 dB SL (squares). The modulation depth of the target was generally adjusted to be just as large as the modulation depth of the reference (medians indicate dm-EMP = 0) and thresholds were not significantly different from 0 (paired t-test). Again, matches were quite accurate, according to the slope and ceiling effect. This is supported by Cronbach's α (α = 0.78).

Parameter	m_{ref}	effect (dB)	shift (%)	slope (%/dB)	ceiling (%)
Bandwidth	0.5	3.65 (1.84)	54.4 (22.9)	16.2 (9.2)	5.50 (6.24)
	0.7	3.05 (1.44)	42.9 (16.0)	19.2 (8.3)	8.00 (9.92)
Center frequency	0.5	2.05 (1.94)	44.2 (22.2)	18.9 (3.0)	1.3 (2.1)
	0.7	0.93 (2.24)	41.7 (33.8)	22.2 (5.9)	0.83 (2.0)
Sensation level	0.5	0.63 (2.7)	49 (21)	25.2 (12.6)	0 (0)
	0.7	0.20 (1.6)	46 (29)	27.2 (6.0)	0 (0)

Table 6.2: *Aspects of behavior for normal hearing subjects determined using the method described in section 6.2.5 using the bandwidth of the stimuli as a parameter.*

6.3.2 Hearing-impaired subjects

Estimates of the parameters for a change in bandwidth are given in Table 6.3. The last two rows gives the averages for normal hearing subjects. As with normal hearing subjects, a positive *effect* was found for most hearing-impaired subjects. Differences between the sizes of the effect for the two reference depths are non-significant for hearing-impaired subjects (paired t-test; p-value = 0.42) as are the differences between normal hearing and hearing-impaired subjects (t-test: p-value = 0.43). For hearing-impaired subjects, there is a significant shift (t-test: p-value < 0.01) towards higher percentages, indicating that most subjects adjust the modulation depth of the target to be more modulated than the reference depth. The *slope*, a parameter indicating the accuracy by which matching experiments are carried out, shows an inconsistent behavior. Relatively small values for HI1 (m=0.5), HI3 and HI4 are reported, whereas others show values close to the values for normal hearing subjects. Differences based on reference depth or hearing capacity are not statistically significant (paired t-test, p-value > 0.1). The ceiling effect was significantly higher for hearing-impaired subjects than for normal hearing subjects (t-test, p-value < 0.001).

Subject	m_{ref}	effect (dB)	shift (%)	slope (%/dB)	ceiling (%)
HI1	0.5	-0.08	83.0	6.5	83
	0.7	0.70	66.5	7.5	66.5
HI2	0.5	1.20	87.5	21.5	12.5
	0.7	2.40	46.0	19.1	12.5
HI3	0.5	7.80	57.5	6.1	30.5
	0.7	7.80	53.5	6.1	30.5
HI4	0.5	5.70	71.0	7.7	41.5
	0.7	4.20	53.0	8.8	33.5
HI5	0.5	1.20	79.0	14.1	12.5
	0.7	1.20	48.0	13.4	16
mean NH	0.5	3.65 (1.84)	54.4 (22.9)	16.2 (9.2)	5.50 (6.24)
	0.7	3.05 (1.44)	42.9 (16.0)	19.2 (8.3)	8.00 (9.92)

Table 6.3: *Identifiers of typical behavior for hearing-impaired subjects determined according to the method described in section 6.2.5 using the bandwidth of the stimuli as a parameter.*

The four parameters that characterize the behavior of subjects in these matching experiments for a change in center frequency are given in Table 6.4. Regarding the size of the *effect*, the results of hearing-impaired subjects are inconclusive; relatively large effects (compared to normal hearing subjects) are reported for subject HI5, opposite effects are reported for subjects HI1, HI2 and HI3, but only for reference depth 0.5 whereas data for HI4 are similar to the data for normal hearing subjects. Data regarding the *shift* are also inconclusive, HI4 and HI5 tended to adjust the modulation depth of the target to be lower than the modulation depth of the reference, whereas other subjects clearly tended to adjust the modulation depth of a target to be higher than the modulation depth of the reference. The *slope* has a comparable discrepancy. Cumulative distributions for HI1, HI4, and HI5 are clearly more shallow than for HI2 and HI3, which are more similar to the data reported for normal hearing subjects. The ceiling effect was significantly larger for hearing-impaired subjects than for normal hearing subjects (t-test, p-value < 0.001).

Table 6.5 gives the results for a change in sensation level given by the four parameters characterizing the behavior of subjects. Hearing-impaired subjects

Subject	m_{ref}	effect (dB)	shift (%)	slope (%/dB)	ceiling (%)
HI1	0.5	-2.55	70	5.8	35
	0.7	2.85	60	6.3	35
HI2	0.5	-0.3	80	10.4	4
	0.7	0.15	65	9.0	10
HI3	0.5	-6.9	70	13.3	34
	0.7	n.a.	n.a.	n.a.	n.a.
HI4	0.5	2.85	60	5.0	20
	0.7	1.95	25	5.6	15
HI5	0.5	6.15	30	6.8	15
	0.7	6.0	35	7.0	10
mean NH	0.5	2.05 (1.94)	44.2 (22.2)	18.9 (3.0)	1.3 (2.1)
	0.7	0.93 (2.24)	41.7 (33.8)	22.2 (5.9)	0.83 (2.0)

Table 6.4: *Similar to Table 6.3, now for a change in center frequency: n.a. stands for not available*

adjust the modulation depth of a target presented at 25 dB SL to be lower than the modulation depth of a reference presented at 10 dB SL. The other way around, the modulation depth of a target presented at 10 dB SL was matched with a higher modulation depth than a reference presented at 25 dB SL. This *effect* on modulation perception induced by a change in sensation level is significantly larger for hearing-impaired subjects than for normal hearing subjects (t-test, p-value < 0.001). Hearing-impaired subjects showed a significant trend to adjust the modulation depth of the target to be higher than the modulation depth of the reference (t-test:p-value < 0.01). The *slope* indicates that most subjects are relatively consistent at adjusting the modulation depth equal to the target. The slope for normal hearing and hearing-impaired subjects is not clearly different (t-test:p-value $= 0.051$). The *ceiling* effect is significantly larger for hearing-impaired subjects than for normal hearing subjects (t-test:p-value < 0.001).

subject	m_{ref}	effect (dB)	shift (%)	slope (%/dB)	ceiling (%)
HI1	0.5	-1.8	60	15.9	35
	0.7	-3.6	65	13.0	45
HI2	0.5	-0.2	100	33.8	10
	0.7	-1.7	50	29.1	5
HI3	0.5	-1.0	85	12.3	40
	0.7	-1.2	75	5.5	40
HI4	0.5	-6.2	70	14.7	25
	0.7	n.a.	n.a.	n.a.	n.a.
HI5	0.5	-5.7	65	22.2	10
	0.7	-5.7	45	22.2	10
mean NH	0.5	0.63 (2.7)	49 (21)	25.2 (12.6)	0 (0)
	0.7	0.20 (1.6)	46 (29)	27.2 (6.0)	0 (0)

Table 6.5: *Similar to Table 6.3, now for a change in sensation level n.a. stands for not available*

6.3.3 Normal hearing subjects with signals presented in noise

In order to simulate the effects of reduced audibility as a result of hearing-impairment for normal hearing subjects, experiments are carried out at similar SPLs and SLs by presenting stimuli in a background noise. The results are given in Table 6.6, with the sensation level for reference and target given by the subscript. The size of the *effect* is not significantly different for normal hearing subjects for signals presented in noise and without noise (paired t-tests; p-value > 0.3). Based on the *shift* of the cumulative distributions, the results of experiments carried out in a masking noise and without a masking noise are significantly different (paired t-tests(8); p-value $= 0.06$). The *slope* is significantly higher for a change in sensation level of 5 dB starting at 10 dB SL ($SL_{10/15}$) (paired t-tests; p-value < 0.05). *Ceiling* effects were not reported for normal hearing subjects without masking noise, whereas the other conditions show ceiling effects in a limited percentage of cases. The reported *effect* size and the *ceiling* effects are still significantly larger for hearing-impaired subjects (paired t-tests; p-value < 0.01).

condition	m_{ref}	effect (dB)	shift (%)	slope (%/dB)	ceiling (%)
no noise	0.5	0.86 (3.34)	56.3 (19.3)	25.2 (12.6)	0 (0)
	0.7	0.30 (1.94)	50.0 (32.4)	27.2 (6.0)	0 (0)
$SL_{10/15}$	0.5	0.60 (0.50)	84.2 (5.8)	17.9 (5.1)	4.1 (4.8)
	0.7	0.86 (1.47)	61.1 (8.8)	22.9 (12.4)	5.4 (10.8)
$SL_{10/25}$	0.5	-0.30 (1.95)	71.5 (11.4)	20.8 (6.7)	3.0 (3.8)
	0.7	0.45 (2.85)	54.7 (8.7)	22.0 (2.7)	5.4 (5.7)
$SL_{25/30}$	0.5	1.43 (1.91)	68.7 (8.0)	19.6 (5.5)	1.0 (2.0)
	0.7	1.16 (1.34)	48.8 (12.0)	30.4 (6.4)	2.1 (2.5)

Table 6.6: *Similar to Table 6.3, now for for four conditions of a change in sensation level averaged for all normal hearing subjects: the four subjects without added noise (no noise), and the conditions with added noise. The sensation levels for reference and target are given by the subscripts*

6.4 Discussion

Normal hearing and hearing-impaired subjects were asked to adjust the modulation depth of a target until it sounded as equally modulated. In each experiment, one parameter was varied at a time (bandwidth smaller or larger than the critical bandwidth (Zwicker, 1961), the center frequency 1 kHz or 4 kHz, or the sensation level 10 or 25 dB SL), while the other two parameters were kept constant. Based on Figure 6.2, the results can be interpreted in terms of effect size (effect), overall shift in perceived modulation depth (shift), accuracy (slope), and ceiling effects (ceiling).

Studying the data, it is striking that most hearing-impaired subjects, independent of the parameters that were varied, adjusted the modulation depth of the target higher than the modulation depth of the reference (shift: dm-EMP > 0). This tendency is unexpected since increasing a parameter should result in an opposite result as decreasing this parameter. Since the target is always preceded by a modulated reference signal, this could indicate that the sensitivity to SAM drops when the stimulus is preceded by another modulated stimulus. In order to reach an equal modulation perception, subjects had to adjust the modulation depth higher than the modulation depth of the target. Similar adaptation to SAM has been reported previously for normal hearing subjects for much longer pre-exposure times (Regan and Tansley, 1979; Tansley and Regan, 1979; Tansley and Suffield, 1983).

In this Discussion two parameters are distinguished, which may have influenced the results described within this chapter;

1. signal characteristics; differences in inherent fluctuations may affect the perception of the target fluctuation (see section 4.1.1)

2. factors of perception; effects such as loudness recruitment, may affect modulation perception by enlarging level differences in the peaks and valleys of the modulation.

6.4.1 Signal characteristics

The total power of inherent fluctuations in the modulation spectrum is independent of the exposure level of a given noise band. Additionally, the envelope spectrum becomes broader and flatter with increasing noise bandwidth (see section 4.1.1). For increasing carrier bandwidths, this leads to a reduction in the envelope power for low modulation frequencies and to an increased envelope power for higher modulation frequencies. In other words, the ratio between target modulation and inherent fluctuations does not depend on the sensation level, but increases for an increasing absolute bandwidth. Within this context, the effects found for normal hearing subjects can be understood. Increasing the bandwidth and center frequency, both result in an increased absolute bandwidth, which reduces the low frequency energy of the inherent fluctuations. As a consequence, the modulation depth of the target is adjusted lower than the modulation depth of the reference (resulting in negative dm-EMPs). Since the ratio of the envelope fluctuations and the target modulation does not change for an altered sensation level, no effect for a change in sensation level on the perceived modulation depth was to be expected.

The absolute bandwidth of the wideband signal with a center frequency of 1 kHz is just as large as the narrowband signal at 4 kHz. Therefore, altering the bandwidth for a 1 kHz signal should result in a similar effect as a change in center frequency for narrowband signals, provided that the bandwidth is the main parameter determining the sensation of the modulation depth and filtering the bandwidth according to the critical bandwidth concept does not alter modulation perception (Eddins, 1999). However, when higher frequencies contribute more to the sensitivity to SAM than lower frequencies (Formby and Muir, 1988), the dm-EMP for an altered center frequency should be larger than the dm-EMP for an altered bandwidth. Figure 6.4 shows a scatter plot of dm-EMPs for a change in center frequency for narrowband signals (ordinate) and a change in bandwidth for the noises with a center frequency of 1 kHz (abscissa). Hence, both panels indicate a similar change in absolute bandwidth (increment on the left side, decrement on the right side). Normal hearing subjects are given by open circles, hearing-impaired subjects by solid diamonds.

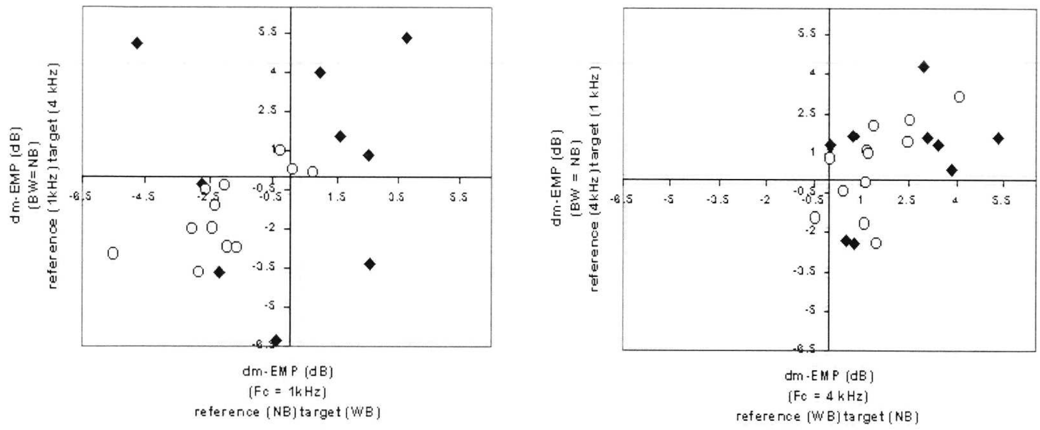


Figure 6.4: *Scatter plots for similar changes in bandwidth. In both cases, the change in absolute bandwidth (Hz) remains the same. Normal hearing subjects are given by open circles, hearing-impaired subjects by solid diamonds.*

Normal hearing subjects showed similar results for the corresponding matching conditions for an altered bandwidth or center frequency (r_{nh}=0.75 [p-value<0.0001]; paired t-test [p-value = 0.6]). This suggests that the perception of amplitude modulations for normal hearing subjects is mainly determined by the absolute bandwidth and thereby the amount of inherent fluctuations and filtering by the basilar membrane does not have an effect on the percept. The results of hearing-impaired subjects are not significantly different from each other, corresponding data however do not result in similar results (r_{hi}=0.18 [p-value > 0.2]; paired t-test [p-value = 0.5]). Apparently, the results of hearing-impaired subjects cannot be explained only in terms of inherent fluctuations. The ceiling effect, shown for hearing impaired subjects, may underestimate the actual correlation between the dm-EMPs for a change in bandwidth and a change in center frequency. Still, hearing-impaired subjects show large differences for the two parameters. This may indicate that the center frequency of the signal alters the perception of the amplitude modulation for hearing-impaired subjects, due to for instance changed signal perception characteristics such as an altered loudness recruitment for the two frequency ranges.

6.4.2 Parameters in perception

If the exposure levels at which signals are presented, are important (see for instance Section 4.3.2), the largest differences would be expected for conditions, at which the center frequency is changed. Table 6.4 indicates that HI4 and HI5 deviate from the other hearing-impaired subjects and normal hearing subjects. Except for having the largest losses, the shapes of the pure-tone audiograms of these subjects are not clearly different from the other audiograms.

It is generally believed that all signals are subjected to a non-linearity which works compressive for normal ears and more linear for impaired ears (see section 4.3.1). This difference in non-linearity, may influence the perceived modulation depth. The slope of the growth of loudness functions* can be seen as a direct psycho-acoustical measure of the magnifying factor by which the compressive non-linearity alters the intensity relations across conditions. These growth of loudness functions may depend upon bandwidth and center frequency since loudness depends on both (Zwicker et al., 1957). In this study, two different situations (different bandwidths, center frequencies, or sensation levels) have been matched in modulation depth to each other. Therefore, the differences in the growth of loudness functions (dgol) may have influenced the differences by which the peak-valley differences of

*The growth of loudness is determined by method of categorical loudness estimation. The procedure, as described by Hellbrück and Moser (1986), divides the dynamic range into 20 equidistant intensity intervals, presented in pseudo-random order. The first stimulus is presented at moderate loudness, and subsequent trials are presented within half the dynamic range from the preceding stimulus. Subjects are provided with nine response intervals ranging from 'very soft' to 'very loud'.
Responses and exposure levels are related to each other using second order functions using the exposure level as the dependent and loudness as the independent variable. The explained variance by these curves is generally high. The derivative of this function is thought to display the growth of loudness.

the modulations of target and reference are magnified relative to each other. In other words, if the modulations in the target signal are enhanced less than the modulations in the reference, the modulation depth of the target must be increased relative to the reference in order to be judged as equally modulated and vice versa. Plasmans et al. (2003) asked subjects to rate the loudness of a signal between 1 and 9 in a study carried out parallel to this study (see footnote). A second order function was plotted relating the loudness of a signal to the actual exposure level, representing the loudness as a function of the intensity. The derivative of this function can be interpreted as the magnifying factor by which small differences in level at a specific sensation level are enhanced: the growth of loudness. The relationship between dgol and dm-EMP has been determined by a correlation coefficient. The correlation coefficients for normal-hearing and hearing-impaired subjects were non-significant for all parameters. The generally small dgols and the relative rough method to determine the growth of loudness by loudness scaling may have obscured the correlation coefficients. For normal hearing subjects, the maximum dgol is typically small (0.054 CLU/dB) compared to hearing-impaired subjects (0.27 CLU/dB). In addition, factors such as the ceiling effect, may have affected the correlation coefficient for hearing-impaired subjects. Moore and Jorasz (1996) showed that the impaired ear enhanced the modulation when compared to the normal ear in unilaterally hearing-impaired subjects, indicating that recruitment has a clear influence on modulation perception.

The effect of growth of loudness is also studied for normal hearing subjects by adding a background noise. The effects found for this simulated hearing loss is in agreement with the trends found for hearing-impaired subjects. However, the effects are more pronounced for hearing-impairment subjects. A small change in sensation level (10 to 15 dB SL) shows a significant difference for simulated hearing loss, for both the shift and slope of the cumulative distributions, compared to normal hearing subjects without added noise. Since hearing-impaired subjects are mainly different in respect to the shift and the ceiling effect, and given the lack of relations with dgol, the effects reported for hearing-impaired subjects may to some extent be determined by lack of audibility, but mostly by other facets of recruitment, such as a loss of compression and, as a consequence of this, a reduction of frequency selectivity.

6.5 Summary & conclusions

The influence of altering one parameter (bandwidth, center frequency or sensation level) on the perception of the modulation depth has been investigated. Results can be summarized as follows:

1. increasing the bandwidth or center frequency for normal hearing subjects results in a modulation depth that is adjusted lower than the modulation depth of the reference. Decreasing the bandwidth or center frequency results in a modulation depth that is adjusted higher than the modulation depth of the reference. Hearing-impaired subjects show a similar trend when corrected for the overall shift in the data.

2. changing the sensation level does not clearly influence modulation perception for normal hearing subjects. Increasing the sensation level for hearing-impaired subjects results in a lower dm-EMP than decreasing the sensation level.

3. the effects for normal hearing subjects can generally be explained by taking the difference in inherent fluctuations into account.

4. differences in dgol did not correlate with the dm-EMPs. This implies that we have not been able to show that recruitment is responsible for the altered modulation perception in hearing-impaired subjects.

5. hearing-impaired subjects generally adjust the modulation depth of the target larger than a given reference. This suggests that the sensitivity to SAM reduces when a signal is preceded by a modulated signal.

The appendix describes the modelling of the data from the normal hearing subjects from the matching experiment using three models: the leaky integrator, the sliding temporal window, and the modulation filter bank. Equal modulation perception is hypothesized to occur on equal statistical output. The leaky integrator and sliding temporal integrator offered generally reasonably high predictive powers. The modulation filter bank showed a lower performance, but did not clearly worsen when the modulation detection threshold is not used as a parameter. In contrast, the predictive power of the leaky integrator and the sliding temporal window clearly worsened when the predictions were only based on the statistical output, without using the sensitivity to SAM.

6.6 Appendix - Modelling the results from SAM-matching for normal hearing subjects

In this Appendix, the data from this chapter will be modelled using three different types of models known to describe temporal processing. These models, the leaky integrator (Viemeister, 1979), the sliding temporal window (Plack et al., 2002), and the modulation filter bank (Dau et al., 1997b), have been previously discussed in section 4.7. In order to predict the data, the modulation detection thresholds were obtained for the same subjects using the same stimuli of the modulation matching experiments. These results will be discussed briefly.

6.6.1 Measuring the sensitivity to SAM

Methods

Modulation detection thresholds were determined using a three interval, three alternative forced choice (3I-3AFC) procedure. The modulation depth was increased after one incorrect response and decreased following three correct responses, thus determining the level of 79.1% correct responses (Levitt, 1971). At the beginning of the run, the signal was fully modulated. The initial step size was 4 dB and was reduced to 2 dB after two reversals. The test ended after 10 reversals. The detection threshold was calculated by taking the mean of the final 6 reversals. Each test was carried out twice. If the thresholds found in the two tests differed by more than 2 dB, a third test was carried out. The modulation detection thresholds, $T(m)_{AFC}$, were calculated using Eq. 6.2, where m denotes the modulation index.

$$T(m)_{AFC} = 20\log_{10}(m) \tag{6.2}$$

Visual feedback was given after each response.

Results

Cronbach's α equaled 0.95, which indicates that the tests are very accurate. In addition, no statistical evidence of a substantial learning effect is found according to the linear mixed effects models and a paired t-tests (p-value > 0.1).

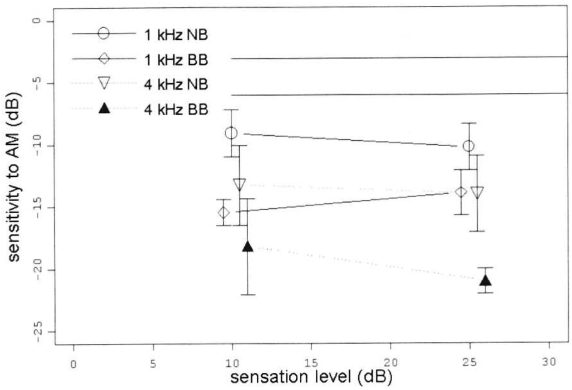


Figure 6.5: *Sensitivity to SAM at two sensation levels for four different signals (see legends upper left corner). All signals were presented at a level of either 10 or 25 dB SL. A small lateral offset was applied to avoid overlapping error bars. Error bars show the standard deviation across subjects. The horizontal lines indicate the reference depths used in the modulation matching experiment.*

Figure 6.5 shows the sensitivity, averaged over subjects, to amplitude modulation as a function of the sensation level. The two horizontal lines indicate the reference depths used in the matching experiment. Main parameters bandwidth and center frequency were both significant (LME, p-value < 0.0001). Increasing these parameters increased the sensitivity to SAM by 5.4 dB for the bandwidth and 4.1 dB for the center frequency. Main parameter sensation level was not statistically significant (LME, p-value $= 0.018$). First order interactions were not statistically significant.

Discussion

Sensitivity to SAM increased with bandwidth, indicating that subjects may be distracted by the intrinsic fluctuations, which are less prominent in wideband signals. Similar results were described by Dau et al. (1999) and these results are in close agreement with their data for a modulation rate of 5 Hz, which is close to the target modulation of this study. Thresholds for a wideband signal at 1 kHz are not significantly different from the thresholds for a narrowband signal at 4 kHz (paired t-test; p-value = 0.25). In addition, there is a trend for subjects to be more sensitive to SAM at higher sensation levels, which is also consistent with literature (Kohlrausch, 1993; Kohlrausch et al., 2000).

The differences between the conditions measured in the two experiments of this chapter are larger for the modulation detection thresholds than for the differences reported by modulation matching. The modulation detection thresholds differ 5.4 dB and 4.1 dB for the main parameter bandwidth and center frequency, respectively. For the modulation matching experiment, these main effects result in differences of 3.1 dB and 1.5 dB. Furthermore, Table 6.2 suggests that these differences decrease for increasing reference depths, as was indicated by a significant main effect for reference depth obtained by pooling the data for the three different parameters (LME, p-value < 0.001). Increasing the reference depth reduced the difference between conditions by 1 dB. Apparently, the perception space for amplitude modulations does not follow the physical rules for modulation depths ($20\log_{10}[m]$), but acts compressive. Similar findings were suggested based on the findings of Chapter 5.

Analogous to the sensation level of a signal, the modulation sensation level can be defined as the amount of decibels the modulation is presented above the threshold of modulation detection. It is apparent that the modulation sensation levels differ for the different conditions presented in Figure 6.6 (distance dashed lines and hatched blocks). The 1 kHz NB-noise signal, for instance, is presented (for reference depth 0.5) at a 3.2 dB modulation sensation level, whereas the 1 kHz BB-signal is presented at a modulation sensation level of 9.3 dB. Hence, the difference in modulation sensation level equals 6.1 dB for a change in bandwidth at 1 kHz. The arrows start from the reference depth pointing to the matched modulation depth for a change in bandwidth with a center frequency of 1 kHz (left panel) or for a center frequency of 4 kHz (right panel). The arrows are parallel for situations going either

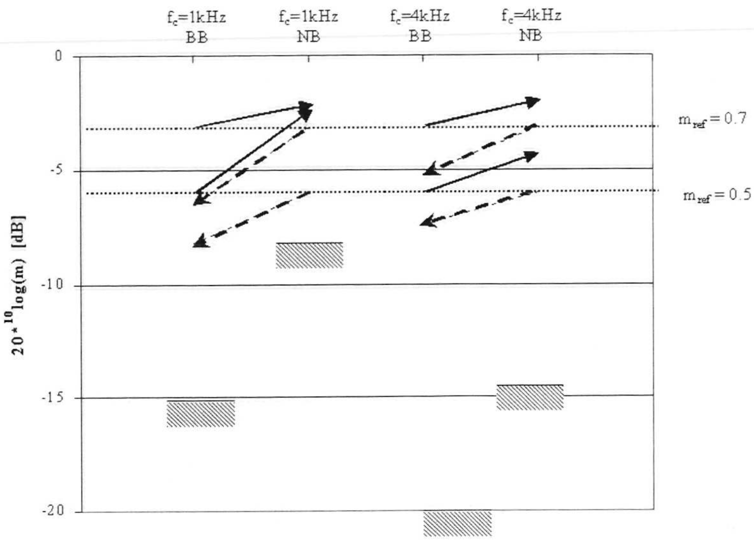


Figure 6.6: *Compression on the perception of SAM. The hatched blocks give the sensitivity to SAM, dashed lines show the two reference depths. Arrows identify the result of the matching experiments.*

down in modulation sensation level or for large dynamic ranges in the modulation depths (4 kHz). However, arrows converge when the bandwidth is decreased, at least for large differences in modulation sensation level. This indicates that modulation perception converges for high modulation depths, which suggest a compression-like system in the modulation domain.

6.6.2 Modelling supra-threshold SAM-perception

Description of the models

Three different models are applied to account for the psychophysical data from the matching experiment. All three models are known to describe temporal processing in the auditory system and have been described more explicitly in section 4.7.

Model 1: The leaky integrator

The leaky integrator (Viemeister, 1979), is primarily developed to describe the sensitivity to SAM as a function of the modulation rate for broadband carriers. A computational version of this model consisting of three† successive stages is implemented: (1) a half-wave rectifier; (2) low-pass filter (first order Butterworth) with a 3-dB cut-off frequency at 64 Hz; (3) a decision device generating a statistical representation of the incoming wave. Five different statistical representations of the decision device have been distinguished and used in this Appendix (see section 4.7, Strickland and Viemeister, 1996).

Equal SAM perception is believed to occur when the statistical outputs are equal according to one of the statistical devices (Equation 4.4-4.8). In other words, the target and reference are expected to be perceived as equally modulated for an equal statistical output. The statistical output is determined as a fraction of the statistical output at the detection threshold to SAM for that particular condition and the statistical output at a particular modulation depth (*e.g.* m = 0.5).

Model 2: The sliding temporal window

This model is described comprehensively by Plack et al. (2002). In the original model, the output of the model is based on the ratio between the signal plus masker and the masker alone. In this Chapter, this decision device is replaced by the same five statistical devices as used for the leaky integrator (model 1; Equation 4.4-4.8). Equal modulation perception is believed to occur for conditions with an equal statistical output according to one of the aforementioned statistical devices for the two conditions in the matching experiment. The statistical output on which equal nodulation perception was determined, is given by the fraction of the statistical output at a given modulation depth and the statistical output at the modulation detection threshold to SAM.

† A 2000 Hz wide pre-detection filter is incorporated in the original model of the leaky integrator, ranging between 4000 and 6000 Hz (first order Butterworth). Since this pre-detection filter is probably not applicable to the conditions described in this Chapter, it was removed from the model. Results, as given in Table 6.7 with and without this pre-detection filter were generally comparable.

Model 3: The modulation filter bank

This model is a simplified version of the model of Dau et al. (1997b). The model consists of four computational stages: (1) a half-wave rectifier (2) followed by an FFT that determined the energy of the target modulation (3) followed by a model output-device according to:

1. the ratio of the energy of the 8 Hz component to the energy falling in the 8 Hz modulation filter. Energy values between 5.5 Hz and 10.5 Hz are summed, following the 5 Hz bandwidth according to Dau's model. Hence, the energy of the target modulation frequency is determined relative to the inherent fluctuations falling in the same modulation filter.
$$\mathrm{MTF}_{Dau_1} = \frac{E_8}{\sum_{5.5}^{10.5} E} \tag{6.3}$$
2. The energy around the modulation frequency of the target, given by the surrounding spikes at 8 Hz, is determined relative to the energy falling in the modulation filter of 8 Hz.
$$\mathrm{MTF}_{Dau_2} = \frac{\sum_{8-df}^{8+df} E}{\sum_{5.5}^{10.5} E} \tag{6.4}$$
3. The energy of the target modulation determined relative to the energy of the signal passing a 2^{nd} order butterworth bandpass filter centered at 8 Hz (approximately 6 Hz wide)
$$\mathrm{E}_{rel_1} = \frac{E_8}{E} \tag{6.5}$$
4. The energy around the target modulation determined relative to the energy of the signal passing a 2^{nd} order butterworth bandpass filter centered at 8 Hz
$$\mathrm{E}_{rel_2} = \frac{\sum_{8-df}^{8+df} E}{E} \tag{6.6}$$
5. The energy of the 8 Hz component itself
$$\mathrm{E}_8 = E(8) \tag{6.7}$$

Again, equal modulation perception is thought to occur based on equal statistical output of the statistical devices. The statistical output is determined relative to the value of the statistical output at the detection threshold to SAM for that particular condition.

Applying the models

Figure 6.7 shows how the level of equal statistical output is determined. As a function of the modulation depth (abscissa), the output based on one of the above mentioned statistical devices (Kurtosis, leaky integrator) is determined. The first panel of Figure 6.7 shows two curves, the dashed curve is based on the narrowband noise with a center frequency at 1 kHz, the solid curve is based on the wideband signal with a center frequency of 1 kHz. The cross and circle give the sensitivity to SAM for these conditions (see section 6.6.1). These values are used in order to normalize the system by dividing all obtained values by the statistical output at the detection threshold, as given in the middle panel. The third panel is a result of enlarging the middle panel. The equal statistical output is represented by arrows. Decreasing the bandwidth using a center frequency of 1 kHz at 10 dB SL, results in dm-EMP(1): a positive dm-EMP according to the data of the modulation matching experiment. Increasing the bandwidth results in a negative dm-EMP given by dm-EMP(2) following the data of the modulation matching experiment[‡].

In the original version, two of the three models are subjected to basilar membrane filtering. In a first order approach, the model output is determined as if signals are not subjective to basilar membrane filtering by skipping this stage, which is reasonable for low modulation rates (*i.e.* 8 Hz).

The compression ratio (see also section 4.3.2 and 4.3.1) may have an effect on the perceived amplitude modulation, by decreasing the perceived modulations. In order to study the effect of the compression ratio, three situations are distinguished: no compression (CR=1.0), an uniform compression ratio of 0.2, and a two-stage compression (CR = 0.78 for low levels followed by a more compressive part for higher levels CR = 0.16). The uniform compression ratio of 0.2 follows the general ideas of compression in the normal hearing ear for medium levels. For lower levels, the compression ratio is more linear (m=1). The two-stage compression, meets these effects by using a more linear behavior at 10 dB SL (CR=0.78) and a more compressive function at 25 dB SL (CR=0.16), which follows the compression ratios as suggested by Plack and Oxenham (1998).

[‡]Apparently the compression-like system for modulation perception is described by the models given by the converging curves in Figure 6.7. In addition, the left pointing arrow is larger than right pointing arrow.

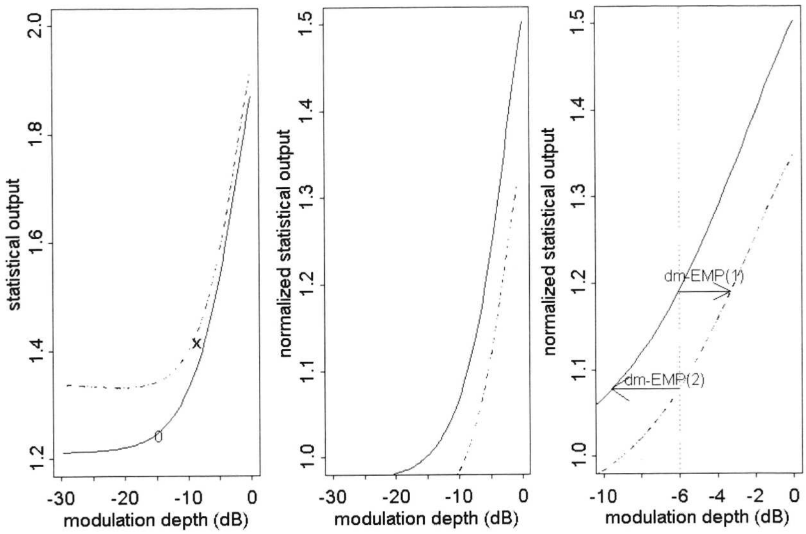


Figure 6.7: *Determination of the equal statistical output for a matching experiment. In the left panel, the statistical fluctuating behavior of a 1 kHz narrowband signal (dashed) and wideband signal (solid) is given as a function of the modulation depth. The 'x' and 'o' reflect the sensitivity to SAM for the two signals (at 10 dB SL). The curves are divided by the threshold for modulation detection (middle panel). The determination of the dm-EMPs for a change in bandwidth for a 1 kHz signal at 10 dB SL is given in the third panel (for m_{ref}=0.5). The arrow dm-EMP(1) illustrates the effect size expected for a decrease in bandwidth using a center frequency of 1 kHz. The arrow dm-EMP(2) illustrates the expected effect of the matching test by increasing the bandwidth.*

Modelling the data

The leaky Integrator

Figure 6.8 shows scatter plots of model predictions (ordinate) and the actual measured dm-EMPs. The left panel shows the results for the leaky integrator using the Kurtosis as a statistical device. Table 6.7 shows the explained variance (R^2) by the leaky integrator for all data together (all; n=48) and for the different parameters that are changed within each matching condition; bandwidth, frequency and level, each based on 16 observations. Model predictions are generally highly significant (** p-value < 0.001; * p-value < 0.01). Using the two-stage compression for two different sensation levels, clearly reduces the performance relative to the linear and the uniformly compressed (CR = 0.2) conditions. Predictive power of the models is generally limited by the poor performance modelling data determined by altering the sensation level[§] (circles). Matching data based on a change in bandwidth (plus-symbols) or center frequency (triangles) are generally well predicted. The models based on the kurtosis and Crest, result in highly significant correlations for both compression conditions (p-value < 0.001). These data suggest that modelling is rather successful. In addition, the data are situated around the dashed line, indicating equality of predicted and measured data[¶].

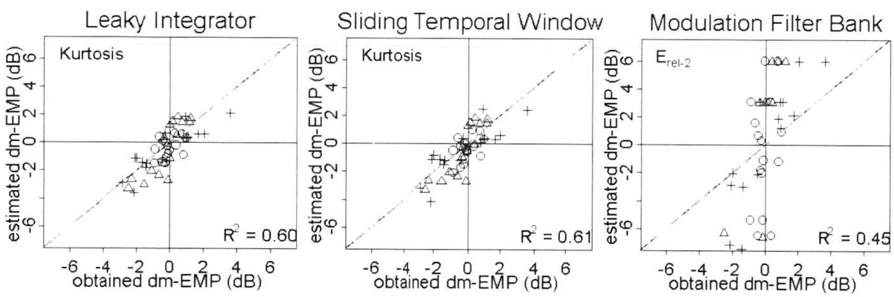


Figure 6.8: *Three examples of model predictions based on models describing temporal processing. The explained variance (R^2) and statistical device. Plus-symbols show the data for matching experiments based on a change in bandwidth, triangles for a change in center frequency and circles a change in sensation level.*

[§] In the original model, Viemeister (1979) used the increase of the (ac-coupled) RMS value of the standard deviation (0.5 dB difference for the signal interval and non-signal interval) as the decision statistic. This is relatively independent of level. The poor predictive power for sensation level is most likely attributable to the relative small differences found for this parameter.

[¶] Several methods may be applicable to determine equal modulation output (see Figure 6.7). Firstly, the statistical output can be determined relative to the statistical output at detection threshold. Secondly, the statistical output can be determined relative to the statistical output at a certain modulation depth (*e.g.* -30 dB), in which case it can be considered as a noise floor determined by signal properties. Thirdly, the statistical output on its own. The big advantage of the last two approaches is that there is no additional information needed (detection threshold). The results of the last two approaches indicate a significant correlation but show a relatively poor predictive power. The statistical output can be determined as a fraction or relative to the statistical output at threshold for SAM detection. Both cases imply that the noise floor, limiting the sensitivity to SAM, is not just given by the inherent fluctuations within the signal, but also by external factors not incorporated within the output device. By subtracting the statistical output at modulation detection threshold, one implies that equal modulation perception occurs at equal statistical output above detection threshold. Whereas dividing the statistical output at detection threshold implies that equal modulation perception occurs at a constant fraction. In this case the fraction was determined since it offered slightly better results.

CR	Parameter	stdev	kurtosis	Crest	maxmin	Slope
1	all	0.59**	0.61**	0.36**	0.28**	0.10
1	frequency	0.72**	0.70**	0.47*	0.28	0.32
1	level	0.26	0.18	0	0.48*	0
1	bandwidth	0.72**	0.77**	0.67**	0.35	0.16
0.2	all	0.1	0.57**	0.53**	0.28**	0.25**
0.2	frequency	0.30	0.72**	0.62**	0.28	0.37
0.2	level	0.08	0.26	0.07	0.48*	0.12
0.2	bandwidth	0.06	0.63**	0.81*	0.35	0.55**
0.78-0.16	all	0.03	0.07	0.14*	0.17*	0.05
0.78-0.16	frequency	0.01	0.80**	0.52*	0.28	0.51*
0.78-0.16	level	0.08	0.1	0.05	0.25	0.11
0.78-0.16	bandwidth	0.75**	0.62**	0.72**	0.35	0

Table 6.7: *Explained variance for the leaky integrator. Each column shows the statistical device which was used in the model. The first column shows the compression ratio that was used. Asterisks give the level of significance (*p-*value<0.01 *; *p-*value<0.001 **).*

The sliding temporal window

Table 6.8 shows the squared correlation coefficients between the predicted values and the obtained values for the matching experiment, based on the sliding temporal window model (corresponding to a low-pass filter with a cut-off frequency of approximately 40 Hz). Model predictions are generally highly significant (** p-value < 0.001; * p-value < 0.01). The predictive power for the linear condition or a compression ratio of 0.2 is generally high. Using different compression ratios for different sensation levels, generally reduces performance. The statistical device influences whether the linear condition leads to a higher predictive power or the uniform compressed condition (CR $= 0.2$). The predictions for experiments in which the bandwidth (plus-symbols) or center frequency (triangles) were altered are generally high, whereas the predictions for experiments in which the sensation level was altered (circles) leads generally to a lower performance. The middle panel of Figure 6.8 shows a scatter plot between measured and predicted dm-EMPs for the sliding temporal window (Kurtosis: CR = 1). Obtained and estimated values are not only closely related, but also nicely situated around the dashed line representing similar values for the model predictions and the measured data[‖].

CR	Parameter	stdev	kurtosis	Crest	maxmin	Slope
1	all	0.07	0.60**	0	0.53**	0.04
1	frequency	0.68**	0.72**	0	0.74**	0.01
1	level	0.24	0.24	0	0.30	0.21
1	bandwidth	0.02	0.70**	0.04	0.61**	0.10
0.2	all	0.50**	0.56**	0.53**	0.54**	0
0.2	frequency	0.70**	0.72**	0.65**	0.70**	0.05
0.2	level	0.26	0.26	0.34	0.44*	0.06
0.2	bandwidth	0.57**	0.69**	0.69**	0.59**	0.01
0.78-0.16	all	0	0	0	0.04	0.01
0.78-0.16	frequency	0.74**	0.83**	0.04	0.69**	0.03
0.78-0.16	level	0	0.02	0.02	0.11	0
0.78-0.16	bandwidth	0.70**	0.66**	0.19	0.50*	0.14

Table 6.8: *Similar to Table 6.7, but now for the sliding temporal window.*

The modulation filter bank

Table 6.9 shows the explained variance of the measured dm-EMP and the predicted dm-EMP values based on the modulation filter bank model. Again, data are shown for three different compression ratio configurations. Using two different compression ratios for signals exposed at 10 and 25 dB SL does generally not improve model

[‖]Similar to the leaky integrator, different approaches of handling the statistical output can be applied. Firstly, the outcome of the statistical output did not result in an accurate prediction of dm-EMP values. Secondly, determining the statistical output relative to the statistical output at -30 dB (the statistical output as determined by the signal properties). In some conditions this leads to highly significant predictions with reasonable predictive power (CR=1; Kurtosis 37%). Thirdly, the statistical output determined as a fraction or relative to the statistical output at threshold for SAM detection. In this case the fraction is determined since it offered slightly better results.

predictions, on the contrary; it generally decreased performance. Generally, similar predictive powers were found for uniform compression and the linear condition. Although the predictive power of the models applied is generally high, the data are not nicely situated around the dashed line. This indicates different dm-EMPs for the model output and the measured data (see right panel Figure 6.8). Again, data for a change in sensation level are poorly predicted**.

CR	Parameter	MTF_{Dau-1}	MTF_{Dau-2}	E_{rel-1}	E_{rel-2}	E_8
1	All	0.02	0.23**	0.16*	0.45**	0.35**
1	frequency	0.097	0.40*	0.51*	0.52*	0.51*
1	level	0.06	0.01	0	0.05	0.06
1	Bandwidth	0	0.36	0.25	0.78**	0.63**
0.2	all	0.06	0.26**	0.37**	0.42**	0.39**
0.2	frequency	0.05	0.69**	0.58**	0.56**	0.72**
0.2	level	0.01	0	0.06	0.04	0.07
0.2	Bandwidth	0.15	0.37	0.60**	0.71**	0.62**
0.78-0.16	all	0.1	0.21*	0.18*	0.36**	0.08
0.78-0.16	frequency	0.31	0.47*	0.50*	0.50*	0.60**
0.78-0.16	level	0.13	0.10	0	0.08	0.01
0.78-0.16	Bandwidth	0.03	0.17	0.59**	0.67**	0.60**

Table 6.9: *Similar to Table 6.7, but now for the modulation filter bank.*

Discussion of the models

Three models are applied and adjusted in order to fit the data from the matching experiment. Results for the different models show a similar behavior in the sense that best results are generally established by using the same compression ratio at 10 and 25 dB SL. The predictive power of the models is generally high, except for the conditions in which the sensation level is changed. This can probably be attributed

**Similar to the leaky integrator, different approaches of handling the statistical output can be applied. Firstly, the outcome of the statistical output device resulted in highly significant predictions with reasonable predictive power (CR=0.2; E_{rel-2} 30.4%). Similar performance could be reached by subtracting the statistical output at a modulation depth of -30 dB. However, subtracting the statistical output at threshold clearly resulted in a higher performance.

to the small effects found for a change in sensation level in the matching experiment (Figure 6.3 bottom). These small effects increase the covariance within the data and introduce noise, which reduces the predictive power. Data given in Table 6.7-6.9 give the explained variance without the correction of attenuation (Equation 5.4). When the correction of attenuation is applied, representing the explained variance when the results of the measured dm-EMPs (matching experiment) and estimated dm-EMPs (indirectly given by the threshold for modulation detection) are not subjective to measurement errors, the 61% given in Table 6.7 becomes actually 75%, due to a multiplication factor of 1.23 (modulation detection α_x=0.95 | modulation matching: α_y=0.7).

For two models, the leaky integrator and the sliding temporal window, model predictions are best when the decision device is determined based on the kurtosis device[††]. In both cases, the predictive power is high and the measured dm-EMP values are similar to the estimated values based on the models. The modulation filter bank is less successful, but is still able to explain 45% (55% after correction of attenuation) of the variance within the data. These results only hold for the data obtained in this experiment and the model uses the sensitivity to SAM as one of the parameters. The model of Dau et al. (1999) is known to predict this sensitivity to SAM as a function of carrier bandwidth quite accurately. The leaky integrator is primarily developed to simulate the TMTF-data for wideband carriers (Viemeister, 1979). The sliding temporal window is known to be less successful in predicting the TMTF, especially for high modulation rates (Moore et al., 1988), which is probably of little importance since 8 Hz was used as the target modulation frequency. The success of the leaky integrator and the sliding temporal window can be attributed to the relative broad noises compared to the cut-off frequency of the low-pass filters used in these models. Both models are unlikely to distinguish between different bandwidths for bandwidths smaller than these cut-off frequencies and will thereby be less successful in predicting matching data using these kind of bandwidths. In these situations, the bandpass filters used in the modulation filter bank, will be able to distinguish between different bandwidths as shown by Dau et al. (1999).

[††] The TMTF for broadband noises was originally modelled using the standard deviation and the max/min as a decision device (Viemeister, 1979). Results from Strickland and Viemeister (1996) indicated that the Kurtosis as a decision device modelled the TMTF with reasonable accuracy.

Chapter 7

The relevance of amplitude modulations to speech perception

In a silent environment, verbal communication presents few problems for normal hearing subjects or subjects with mild or moderate losses. Speech presented simultaneously with other speakers or subjected to reverberation, causes difficulties, especially for hearing-impaired subjects. The reason for the reduced performance of hearing-impaired subjects with regard to normal hearing subjects is primarily related to a reduced audibility. Since thresholds are elevated, speech is partly inaudible. However, at least part of the problem can be attributed to additional difficulties in segregating speech signals from background noise, due to an impaired temporal or spectral resolution (see Chapter 4).

This Chapter describes results, which illustrate the importance of the envelope to the intelligibility of speech. An important aspect is the ability of normal hearing subjects to improve speech intelligibility when modulated noise is present instead of continuous noise at the same signal-to-noise ratio. Hearing-impaired subjects obtain less or no benefit from the fluctuations in the masking noise. In the remainder of this thesis, the extent to which perception of speech by hearing-impaired subjects is limited by a reduced sensitivity to SAM is examined. Since speech is a highly modulated signal. speech perception may be affected by MDI when speech is presented together with modulated maskers. such as fluctuating noise. Reduced masking release for speech (MRS) may be caused by excessive modulation masking.

7.1 Speech in noise

Both. noise and hearing-impairment. may reduce audibility. Intelligibility can be predicted by dividing the spectrum of speech into a number of bands. The signal-to-noise (S/N) ratio. or reduction in audibility due to hearing loss is then determined for each band. Weighing each band by a factor representing the importance of the band to overall intelligibility* and summing, results in the speech intelligibility index (SII). The SII ranges from 0 (not intelligible) to 1 (completely intelligible), giving the proportion of speech that is still available to the listener. Depending on the weights used. the SII can predict speech intelligibility for linear distortions such as a reduced audibility due to hearing-impairment or added noise. However. Pavlovic (1984) indicated that a model predicting intelligibility based on the audibility of the speech spectrum did not correctly predict intelligibility for subjects with moderate to severe hearing losses. This implies that factors, other than reduced audibility, may degrade the ability to understand speech. Noordhoek et al. (2000) showed that for subjects with hearing losses larger than 25 dB SL other parameters than audibility must limit the ability to understand speech. However. some subjects with larger losses may have 'normal' SII values. These subjects' disability can be attributed to audibility alone.

7.1.1 The Speech Reception Threshold (SRT)

A well accepted measure of speech intelligibility is given by the speech reception threshold (abbreviated as the SRT: Plomp and Mimpen. 1979). The SRT is defined as the S/N ratio in decibels at which 50% of the sentences are repeated correctly. Performance strongly depends on the S/N ratio. Normal hearing subjects are unable to repeat sentences correctly for a broadband noise with the same long term average spectrum as the speech at a S/N ratio of approximately -8 dB. whereas all sentences

*Weighting factors and calculation routines are defined for octave bands. 1/3 octave bands. and critical bands.

are repeated correctly at a S/N ratio of -1 dB (Festen and Plomp, 1990; see open squares Figure 7.1). Hence, the range from completely intelligible to completely unintelligible is 7 dB. The SRT is about -5 dB. Hearing-impaired subjects generally perform worse than normal hearing subjects. This leads to higher SRT-values, ranging from -2 dB for subjects with mild or moderate losses to 10 dB for subjects with more severe losses. The range of S/N-ratios at which speech is completely intelligible and at which speech is not intelligible is comparable to what was reported for normal hearing subjects (see solid squares Figure 7.1). Although an elevation in SRT-score by 2-3 dB appears to be modest, it is sufficient to create a substantial loss of intelligibility in difficult situations.

7.1.2 Speech in fluctuating noise

Speech intelligibility is usually better for normal hearing subjects when the speech is presented in a fluctuating background noise (triangles) rather than a continuous background (squares) noise at the same S/N-ratio (Festen and Plomp, 1990; Howard-Jones and Rosen, 1993a; see Figure 7.1). When compared to normal hearing subjects (open symbols), hearing-impaired subjects (solid symbols) experience reduced intelligibility in continuous noise (squares). In addition, intelligibility does not increase when fluctuating noise is presented at the same S/N-ratio for hearing-impaired subjects.

Subjects benefited from the valleys in the masker leading to temporary improvements in S/N ratio. This improvement in speech intelligibility resulting from fluctuations in the background noise is referred to as *masking release for speech* (MRS) within the context of this thesis. The release of masking is largest for block-wave modulated noise at approximately 10 Hz (Miller and Licklider, 1950). In another study by Festen (1987), the SRTs for normal hearing and hearing-impaired subjects were obtained in a continuous background noise and a sinusoidally *intensity* modulated noise modulated by 4, 8, 16, and 32 Hz, having the same long term average spectrum as speech. For normal hearing subjects, the smallest masking release for speech (3 dB) is obtained for noise modulated by 4 Hz and the largest masking release for speech is found for modulation rates of 16 and 32 Hz (5.5 dB). These results presumably represent a trade off between the masking of complete words at low modulation rates and insufficient temporal resolution at higher modulation

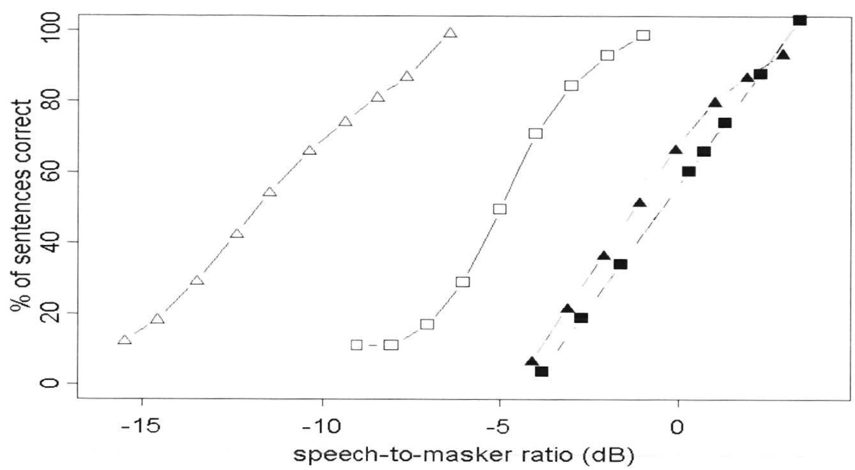


Figure 7.1: *Average discrimination curves for sentences presented in steady state noise (squares) and fluctuating noise (triangles), for normal hearing (open symbols) and hearing-impaired listeners (solid symbols) (based on 140 responses)*
redrawn from Festen and Plomp (1990)

rates. However, for hearing-impaired subjects, the SRT is 3.2 dB higher than the SRT in continuous noise. This is attributed to the reduced temporal resolution of the hearing-impaired subjects relative to normal hearing subjects. Trine (1995) measured the effect of Gaussian noise, interrupted at rates of 2, 4, 8, 16, 32, 64 and 128 Hz, on intelligibility for normal-hearing subjects. Subjects were asked to adjust the level of the noise until they thought that 50% of the speech was intelligible. Masking release for speech was approximately 22 dB for modulation rates below 16 Hz and decreased monotonically to 0 dB for a rate of 128 Hz. Gustafsson and Arlinger (1994) determined that, by using SAM noise as a masker, higher modulation ($>$ 10 Hz) rates produced less masking release than lower modulation rates and modulation depths of -6 dB ($20\log_{10}(m)$) resulted in a smaller release of masking than modulation depths of -12 dB or 0 dB (100%).

The most obvious example of a fluctuating background noise is given by a competing speaker†. Hearing-impaired subjects often report great difficulties understanding speech with one interfering talker. The masking release for speech found for normal hearing subjects (difference in squares and triangles) is approximately 7 dB. For hearing-impaired subjects, SRTs for a fluctuating and a continuous background noise usually are comparable. Hence, the difference in the SRT using a fluctuating background noise can be over 12 dB.

The benefit of the fluctuations in noise can only partly be attributed to a reduced audibility. Bacon et al. (1998) measured speech intelligibility for normal hearing subjects in a fluctuating noise to which a continuous background noise was added to simulate the reduced audibility of hearing-impaired subjects. The masking release for speech reduced, but fluctuations in the background noise were still beneficial to normal hearing subjects. This implies that the reduced intelligibility in a fluctuating noise in hearing-impaired listeners is at least partly due to supra-threshold deficits.

7.2 Factors limiting the intelligibility in noise

The supra-threshold deficits limiting speech intelligibility can be attributed to the altered spectral and temporal resolution of hearing-impaired subjects (Noordhoek et al., 2001). The effect of hearing-impairment on these measures are discussed in Chapter 4 of this thesis. Two methods can be used to study the effects of these psychoacoustical factors on speech intelligibility :

1. **Correlational studies.** The correlation between the performance on an auditory task and speech intelligibility is determined.

2. **Disturbing the speech signal.** Disturbing the speech signal may provide information on the cues that are important to speech intelligibility. This information can be obtained by:

†The fluctuating noise used to obtain the data in Figure 7.1. is created by filtering concatenated sentences at 1 kHz and replacing the fine structure of the low and high-pass segments with a continuous noise leaving the temporal envelope intact.

(a) determining the effect of specific distortions, such as reverberation, on intelligibility by leaving the speech signal intact[‡].

(b) studying the effect of reducing information within the speech signal on intelligibility by, for instance, filtering.

7.2.1 Correlational studies

A variety of studies have examined the relationship between the intelligibility of speech presented in noise and a diversity of spectral and temporal psychoacoustical parameters.

Festen and Plomp (1983) indicated that measures of spectral resolution, such as critical ratio and the low frequency edge of the estimated PTC, are clearly related to the intelligibility of speech in noise. These findings have been confirmed using parameters such as the critical ratio (Dreschler, 1983), the width of the psychoacoustical tuning curve (Horst, 1987) and just noticeable differences in frequency (Glasberg and Moore, 1989). Festen and Plomp (1983) determined speech intelligibility related to measures of temporal resolution, such as forward masking, backward masking and click thresholds in noise. Relatively weak correlations were reported. This has been confirmed by other studies and for other measures of temporal resolution, such as the sensitivity to SAM (Takahashi and Bacon, 1992) and forward and backward masking (Dreschler, 1983). However, results for gap-detection are inconclusive. Modest correlation are reported for speech in noise (Dreschler, 1983) or relative to the STI in a reverberant room (Dreschler and Leeuw, 1990). However, strong effects have been reported by Glasberg and Moore (1989).

Speech intelligibility with correction of audibility (SII) was measured with regard to auditory performance for tasks measuring temporal or spectral resolution by Noordhoek et al. (2001). Positive significant correlation coefficients were reported, suggesting that supra-threshold deficits indeed limit speech intelligibility. The components of a PCA on the psychoacoustical measures indicated that three components, temporal resolution (shape discrimination, forward and backward masking), spectral discrimination (frequency discrimination), and spectral resolution

[‡]Applying distortions to a signal will generally alter the signal within certain limits. For instance, applying a gain to certain frequency components will alter the temporal structure and phase relations.

(upward and downward spread of masking) contributed significantly (p-value < 0.05) to speech intelligibility in noise.

CMR related to the intelligibility of speech

Co-modulation Masking Release (CMR) refers to the increased ability to detect a signal in noise using information on the envelope of the masker outside the filter tuned to the target. Howard-Jones and Rosen (1993b) measured the effect of 'uncomodulated glimpsing' on masking release using a 'checkerboard' noise[§]. The masking release for a fully co-modulated background noise is approximately 23 dB and decreases when multiple bands are modulated in anti-phase and is absent for 8 or more spectral bands. When only two bands are used, the 'checkerboard' noise results in a higher masking release than when the noise contains one modulated channel and one continuous channel. Apparently, listeners benefit from 'uncomodulated glimpsing' by combining speech cues from different frequency bands at different times in the signal. In section 4.4.2, literature has been discussed that indicated that CMR vanished when masker envelopes modulated 180^0 out of phase were used. In addition, using speech as a target signal introduces two differences with traditional CMR (Grose and Hall, 1992; see section 4.4):

1. the target (speech) is a complex broadband signal rather than a spectrally discrete signal (tonal).

2. the tasks (speech intelligibility) involves supra-threshold recognition rather than detection.

Festen (1987) used noise divided into two 100% SIM[¶] modulated bands running in and out of phase. No clear benefit for co-modulation at 4 Hz was reported and a relatively small benefit of approximately 1 dB at the highest modulation rate (32 Hz). Grose and Hall (1992) reported a clear effect of CMR in the ability to detect speech, whereas no clear effect of CMR on intelligibility was found. Festen (1993) measured speech intelligibility by masking speech with speech divided into multiple

[§]broadband noise was split into multiple bands (2, 4, 8, or 16) of equal power and modulated either in phase or out of phase at a 10 Hz rate. A spectogram of this type of signal results in an image that looks like a checkerboard

[¶]Festen used sinusoidally *intensity* modulated noise, in order to approach the effect of an interfering speaker more realisticly.

bands. The effect of CMR, determined by the difference in intelligibility for speech presented in a masking voice with and without time delay between 1/3 octave bands, was small. Again, the phases of the envelopes are crucial to obtain a fair amount of CMR, and the weak relationship between CMR and masking release for speech may not indicate that the relationship does not exist. Kwon (2002) measured the recognition of band-filtered consonants masked by multiple bands and observed a small effect (3.5%) for CMR.

CMR can be obtained for signal detection tasks (Hall and Grose, 1988) or the detection of speech (Grose and Hall, 1992). However, the literature is inconclusive as to whether CMR occurs for supra-threshold recognition. For tasks such as gap detection, conditions benefitting CMR increase the sensitivity to the gap (Hall and Grose, 1992). In contrast, the sensitivity to other supra-threshold recognition tasks such as speech intelligibility (Grose and Hall, 1992; Festen, 1993; Kwon, 2002) and amplitude modulation discrimination (Hall and Grose, 1995) did not increase using conditions that enhance CMR.

MDI related to the intelligibility of speech

A limited number of studies have focused on the intelligibility of speech in relation to measures as MDI. This is unsurprising since the intelligibility of speech improves, while the sensitivity to SAM is reduced when modulaters maskers are used. However, MDI may reduce the masking release for speech. Takahashi and Bacon (1992) reported weak correlation coefficients between modulation detection and masking release for speech, expressed in percentage correct at a S/N ratio of 0 dB. Kwon and Turner (2001) measured the intelligibility of filtered consonants presented in fluctuating masker bands at adjacent frequencies. When the speech and masker were widely separated in frequency, the effect of MDI was dominant.

MDI and masking release for speech : results from Chapter 5

Binaural speech intelligibility for the subjects, who participated in the MDI experiments described in Chapter 5, was determined using a standard SRT-test. Figure 7.2 illustrates the results of the SRT in a fluctuating noise (ordinate) as a function of the SRT in a continuous background noise (abscissa). The diagonal line indicates where the points would fall if identical results were obtained for speech in

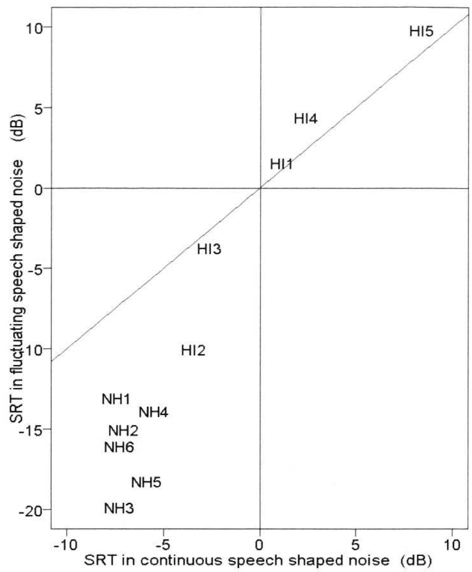


Figure 7.2: *Scatter plot of individual speech intelligibility performances in a continuous (abscissa) and a fluctuating background noise (ordinate). The diagonal line indicates the points for which the SRT in a fluctuating noise equals the SRT in a continuous noise (masking release for speech = 0).*

a continuous and in a fluctuating noise. Hence, masking release for speech is given by the deviation from the diagonal line and is larger for normal hearing than for hearing-impaired subjects.

Since the number of subjects is small, rank-order correlation coefficients were determined (Spearman's ρ) in order to determine whether masking release for speech and MDI are linked. The results suggest that there is no relationship between MDI and masking release for speech (see Table 7.1). The only exception is the reduced sensitivity to SAM with regard to the conditions without maskers (MDI_{PA}) using a reference depth of 0.30 (p-value < 0.05). Figure 7.3 gives the scatter plot between these two parameters. The figure indicates that more interference coincides with a larger MRS, which is in contradiction with the stated hypothesis. Apparently, the reduction in MRS is not related to an excessive amount of modulation masking.

However, determining the Spearman rank correlation coefficients for normal hearing subjects and hearing-impaired subjects separately, indicates that both groups have a negative non-significant correlation coefficient for reference depth 0.18.

	ρ (MRS)
$MDI_{UNM}(0.18)$	0.07
$MDI_{UNM}(0.30)$	0.16
$MDI_{PA}(0.18)$	0.56
$MDI_{PA}(0.30)$	0.62*
MDI_{UNM}	-0.14
MDI_{PA}	0.24

Table 7.1: *Rank correlation coefficients between the masking release for speech (MRS) and the amount of interference (MDI). The masking release for speech was determined using the difference in the SRT for speech in a fluctuating and a continuous background noise. The data between brackets give the reference depth. MDI_{UNM} gives the reduced sensitivity to SAM by adding modulated maskers instead of non-modulated maskers. MDI_{PA} is determined relative to the condition without maskers. Asterisks denote the level of significance,* p-value*<0.05 *;* p-value*<0.01 **.*

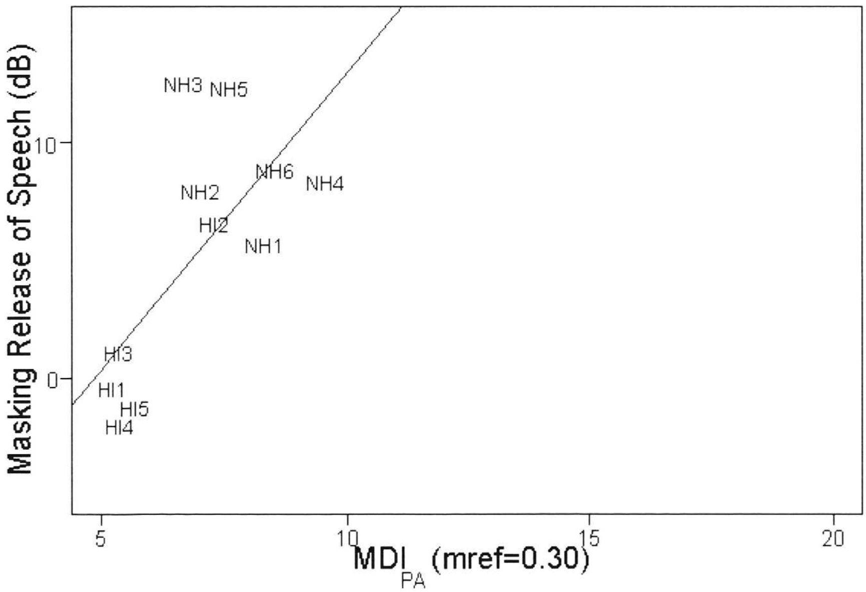


Figure 7.3: *Scatter plot of the release of masking for speech (ordinate) against the interference caused by the modulated maskers at a reference depth of 0.30 (abscissa). The solid line shows the regression line between the data.*

This indicates that there is a tendency for masking release for speech to decrease as MDI increases. However, for reference depth 0.30, both groups exhibit positive correlation coefficients. All correlation coefficients indicate non-significant relations.

	MDI (0.18)	MDI (0.30)
ρ	0.07	0.16
ρ_{NH}	-0.54	0.18
ρ_{HI}	-0.15	0.28

Table 7.2: *Rank correlation coefficients between the masking release of speech and the amount of interference.*

More on the use of fluctuations in the masking signal

The SRT-test is a valuable method for determining the subjects' intelligibility. A small disadvantage of the method is that it does not provide a good insight into the relationship between intelligibility scores (percentage correct) and the S/N ratio used. These curves are given by Festen and Plomp (1990). The performance in percentage is given for normal hearing and hearing-impaired subjects as a function of the S/N ratio of speech presented in a continuous and fluctuating noise (see also Figure 7.1). The results indicate that normal hearing subjects experience improved intelligibility when the masking signal contains gaps, and the benefit of these gaps increases as percentage correct-scores decrease (the curve for speech in fluctuating noise is more shallow than the curve for continuous noise). The gaps in the noise do not improve speech intelligibility for hearing-impaired subjects. In addition, the slopes for the continuous noise and the fluctuating noise are comparable to each other.

In order to assess the ability to use the dips of the fluctuations in a background noise ('glimpsing'), matching experiments have been conducted (not described in this thesis) in which subjects are asked to adjust the level of the speech until speech was:

1. just inaudible [*net niet hoorbaar*]
2. audible [*hoorbaar*]

3. well audible, not intelligible [*goed hoorbaar, niet verstaanbaar*]

4. just intelligible [*net verstaanbaar*]

5. very intelligible [*goed verstaanbaar*]

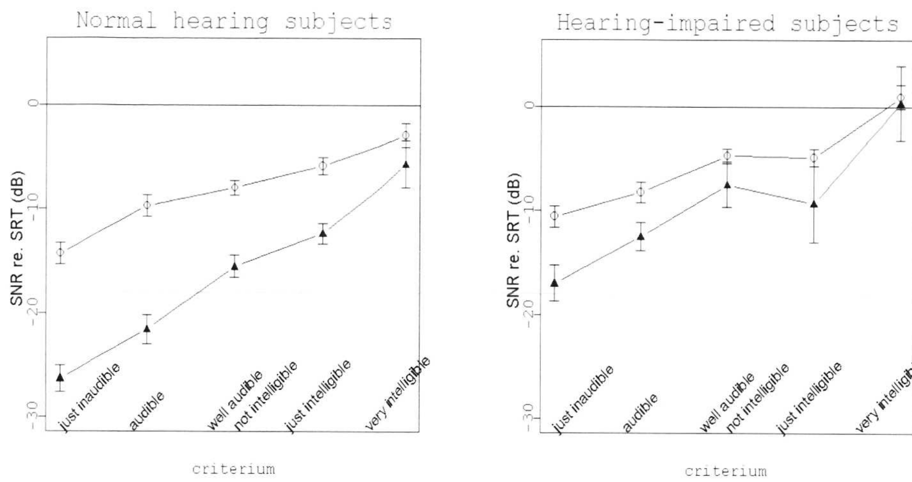


Figure 7.4: *Results of matching experiments carried out for normal hearing subjects (9 subjects; left panel) and hearing-impaired subjects (3 subjects; right panel) carried out in a fluctuating noise (triangles) and a continuous background noise (circles)*

Experiments were carried out in a continuous and in a fluctuating speech-shaped noise in order to determine the masking release for speech. Figure 7.4 shows the mean of the differences for the subjects' S/N-ratio and their SRT for that noise. Hence, the data have been corrected for masking release for speech as determined by the SRT measurements and the difference between the two curves represents the additional benefit of fluctuations. The results clearly indicate that the fluctuating background noise provides additional masking release for the detection of speech for normal and hearing-impaired subjects (diverging curves). As aforementioned, CMR effects have been reported for speech detection, but not for speech intelligibility (Grose and Hall, 1992). However, the monotonic converging curves in Figure 7.4 illustrate that the difference between the intelligibility and the ability to detect speech does not show a sudden increase in masking release for speech.

Given the importance of the coherence in the envelope to CMR, it seems plausible to expect a similar importance of coherence for the target signal. Although the envelope of speech is roughly coherent over frequency bands, Figure 7.5 clearly indicates that, for a randomly chosen sentence, different bands within speech are not completely coherent. There are large contrasts in the amount of energy in the periods indicated by the vertical lines in the different frequency bands. CMR is known to decrease as phase differences for across-band information (the masker) increase. Therefore it is questionable whether CMR can account for the improvement in the intelligibility of speech due to the fluctuating behavior of the masker.

Figure 7.5: *A Dutch sentence filtered into octave bands of 0.5, 1, 2, or 4 kHz. Vertical lines are used to indicate periods in which there is little coherence between bands.*

7.2.2 Disturbing the speech signal

Interference of the intact speech signal

The relationship between speech intelligibility and a distorted auditory coding can be determined using the distortion sensitivity approach (Plomp, 1986). Removing cues that cannot be used due to a distorted auditory coding will not affect speech intelligibility for hearing-impaired subjects. Measuring intelligibility as a function of the distortion applied to the signal will reduce the difference in performance for normal hearing and hearing-impaired subjects. However, when the difference in intelligibility scores for normal hearing and hearing-impaired subjects is constant, independently of the applied distortion, hearing-impaired subjects are as sensitive to the distortion as normal hearing subjects. This indicates that the distortion is not related to the deficits limiting intelligibility of the individual hearing-impaired subject.

Duquesnoy and Plomp (1980) determined the effect of reverberation on speech intelligibility as a function of the reverberation time. Their results indicate that hearing-impaired subjects are as sensitive as normal hearing subjects to reverberation. As a tool to describe auditory processing van Schijndel (2000) used wavelet coding[‖]. A distortion sensitivity model was used to determine the effect of temporal, spectral and level distortions on speech intelligibility. Results indicate that hearing-impaired subjects are as sensitive as normal hearing subjects to a distorted coding of temporal or level information. This suggests that hearing-impaired subjects do not experience additional problems as a result of an altered coding strategy with regard to temporal and level distortion. Applying spectral distortion, clearly indicated that speech perception for hearing-impaired subjects was reduced less than for normal hearing subjects. This indicates that the problems hearing-impaired subjects experience in the intelligibility of speech are mainly caused by sub-optimal spectral coding properties.

To investigate the relative importance of envelope and fine structure, Smith et al. (2002) generated signals that have the envelope of one sound and replaced the original fine structure by the fine structure of the other. The envelope appeared to be most important to speech intelligibility, whereas the fine structure is most important for pitch perception (melody recognition) and sound localization.

Reducing information in the speech signal (spectral)

In section 7.1 the effect of adding noise has been discussed in terms of the SII. The SII assumes that overall intelligibility is determined by the S/N ratio within each

[‖] Wavelet coding is a tool to describe the temporal-spectral behavior of the peripheral part of the ear. To some extent, it can be compared to the short-time Fourier analysis. However, the spectral resolution for wavelet coding decreases with frequency (constant on a log-frequency scale). This is more similar to the coding strategy of the auditory system. Experiments also suggested that the duration of the time-frequency window decreases with frequency.

band, with each band having its own importance weighting. The band importance functions were introduced to account for the effect that some bands contribute more to overall intelligibility than other bands (Pavlovic, 1987). Intelligibility is determined for each band independently, with 1/3 octave bands between 800 Hz and 4 kHz having the highest weights. These frequencies represent 80% of the information in speech. Lippman (1996) indicated that reasonable intelligibility was obtained for CVC-words for frequencies below 800 Hz (44%). However, spectral information above 8 kHz, still offers a significant increase in CVC-score of approximately 30% in intelligibility. Hence, speech is a multi-channel process for which different channels fuse and lead to a single percept. Such interaction is also reported for across-channel processes such as CMR and MDI.

Reducing information in the speech signal (temporal)

Shannon et al. (1995) studied the extent to which fine structure information is essential for speech intelligibility. The original fine structure of speech was replaced by a white noise, while maintaining the envelope information for each band. The number of bands ranged from one to four** . The intelligibility of consonants, vowels, and words is above 80% using three bands. Using four bands increased performance to about 90%, which indicates that nearly perfect intelligibility can be obtained using the temporal structure of just four bands. Replacing the carrier signal of speech by pure tones modulated by the envelope of the band centered at the signal frequency still results in a fair intelligibility (Dorman et al., 1997).

Disturbance of the temporal structure whilst keeping the fine structure intact has been studied by Drullman et al. (1994a,b). The speech signal is divided into spectral frequency bands. The envelope of each band was low-pass or high-pass filtered. Filtering out modulation frequencies *above* 16 Hz barely affected the intelligibility of consonants, vowels and sentences. Filtering out modulation frequencies below 16 Hz gradually reduced speech intelligibility (Drullman et al., 1994b). Filtering out modulation frequencies *below* 4 Hz also barely affected the intelligibility of consonants, vowels and sentences. Filtering out frequencies above 4 Hz reduced the intelligibility of speech gradually (Drullman et al., 1994a). When

**All conditions were low-pass filtered at 4 kHz. Filter cutoff frequencies were 1500 Hz for the two-band processor, 800 and 1500 Hz for the three-band processor, and 800, 1500, and 2500 Hz for the four-band processor.

the two studies are compared, they indicate that the effect of low-pass and high-pass filtering of the envelope is approximately the same at 8–10 Hz. A combination of previous studies has been studied by van der Horst et al. (1999). The consonants of VCV syllables were split into bands of which the original fine structure was replaced by a speech-shaped noise. For each band, modulation frequencies were filtered out around 8, 12 and 16 Hz. A large reduction was found for relatively wide notches ($>$ 10 Hz) around 8 Hz, whereas other conditions result in a marginally decreased performance.

The importance of temporal troughs was studied by Drullman (1995). The troughs were flattened, simulating the effect of added noise. The SRT was 5 to 6 dB better for manipulated speech than for speech in noise, which indicates that the added noise may confuse the listener, for instance by the inherent fluctuations of the added noise. Replacing the fine structure of speech by the fine structure of noise, leaving the temporal speech envelope intact indicates that speech intelligibility is barely affected by manipulations of the fine structure. However, maintaining the fine structure and randomizing the envelope causes intelligibility to drop to 17 %.

Another aspect of the temporal structure of speech was investigated by Greenberg et al. (1998). Speech is divided into one-third octave bands and intelligibility was measured using these bands alone or in combination with each other. Intelligibility was high, even with only three bands. However, desynchronizing these bands by more than 25 ms, severely degraded speech intelligibility. Increasing the number of bands to 19, made listeners relatively insensitive to temporal asynchrony (Arai and Greenberg, 1998).

7.3 Implications

This chapter has showed that:

1. masking release for speech is substantial for normal hearing subjects and absent for hearing-impaired subjects.

2. the temporal envelope of spectral bands in speech is an important factor in the intelligibility of speech.

3. the intelligibility of sentences is high when the temporal envelope of speech

consists of four bands or more.

4. speech is a very robust signal and considerable temporal distortion is needed to reduce speech intelligibility.

Hence, speech is a highly redundant signal, in which perceptual trade-offs occur. Information across auditory filters provides additional cues, which improve speech intelligibility. Given the importance of the envelope to the intelligibility of speech, the results described in section 7.2.1 are surprising.

The lack of evidence supporting a relationship between MDI and MRS, may be due to the difference in gating. In the MDI-experiment, target and masker were gated synchronously, whereas speech intelligibility was assessed using asynchronous gating of speech and masker. Previous studies have indicated that MDI is largest when the probe and masker carriers are gated synchronously, with the amount of interference falling as gating asynchrony is increased (Hall and Grose, 1991; Mendoza et al., 1995b). However, even for continuous gating, a fair amount of MDI was measured. This difference in gating may contribute to the low correlation coefficients. The results described in Chapter 5 indicated large differences between normal hearing and hearing-impaired subjects in the tendency to adjust the signal with a larger modulation depth than the reference depth. This tendency may reflect a reduction in sensitivity to SAM due to pre-exposure to SAM. Speech intelligibility may be affected and the effect may be considered as 'true' MDI, indicating that a reduced sensitivity to SAM also results from factors other then perceptual grouping.

Chapter 8

Masking release for speech related to adaptation to SAM*

Abstract

Modulation detection thresholds were measured for a 1 kHz pure tone preceded by a 1 kHz adaptor that was either non-modulated or fully modulated. The duration of the signals was 625 ms and the inter-stimulus intervals varied between 10 and 625 ms. Signals were amplitude modulated by an 8 Hz sinusoid. Ten normal hearing and twelve hearing-impaired subjects participated in the experiments. Clear differences in sensitivity to SAM were found for hearing-impaired listeners when a modulated adaptor rather than a non-modulated adaptor preceded the target. This indicated that hearing-impaired subjects also show modulation detection interference for non-simultaneous signals. The adaptation effect observed correlated strongly with the release of masking for speech when a fluctuating background noise was used instead of a continuous noise with the same long-term RMS levels.

*submitted to the Journal of the Acoustical Society of America

8.1 Motivation

Chapter 6 describes matching experiments carried out for normal hearing and hearing-impaired subjects. When matching the modulation depth to a reference depth, hearing-impaired subjects tended to adjust the modulation depth of the target to be higher than the modulation depth of the reference. Normal hearing subjects do not tend to demonstrate this tendency. In the Discussion of that Chapter, it was suggested that this could be attributed to some kind of adaptation to SAM. The literature described adaptation to SAM for pre-exposure times of several seconds but not for relatively short adaptation periods as used in Chapter 6 (see also section 4.8). The importance of the temporal envelope for speech perception was discussed extensively in Chapter 7. One might expect that the reduced masking release for speech for hearing-impaired subjects may, in some way, be limited by the increased adaptation to SAM. More explicitly, the inability of hearing-impaired subjects to improve speech intelligibility when speech is presented in a fluctuating rather than a continuous noise with the same RMS-level may be caused by the adaptation to fluctuations in the modulated noise.

This paper describes three experiments. The first experiment measured the sensitivity to SAM of a stimulus without an adaptor. The second experiment measured the sensitivity to SAM of a stimulus preceded by a modulated or non-modulated masker. The third experiment measures the masking release for speech using continuous and fluctuating background noises. The relationship between the sensitivity to SAM and masking release for speech will be examined.

8.2 Methods

8.2.1 Subjects

Ten normal hearing and twelve hearing-impaired subjects participated in this study. The normal hearing subjects had thresholds smaller than 15 dB HL at the standard audiometric frequencies from 125 Hz to 8 kHz (re. ANSI, 1996). The average age was 29 years and subjects were tested at a randomly assigned ear. Table 8.1 represents the age, tested ear and the auditory thresholds of the hearing-impaired subjects. The bottom row of Table 8.1 represents the average age and thresholds of the hearing impaired subjects, with their standard deviations. The subject number increases with the average loss at the frequencies 1, 2, and 4 kHz. The best ear was tested in case of an asymmetrical hearing loss.

			Pure tone thresholds					
Subject	Age	Ear	0.25 kHz	0.5 kHz	1 kHz	2 kHz	4kHz	8 kHz
HI1	43	L	45	45	30	25	40	35
HI2	18	R	30	40	35	35	40	35
HI3	54	L	35	35	40	40	35	25
HI4	59	L	30	35	40	40	40	55
HI5	64	R	45	45	40	30	60	75
HI6	51	R	25	25	35	40	60	60
HI7	62	L	30	45	50	50	50	40
HI8	62	L	30	40	45	50	60	65
HI9	59	L	15	10	45	60	55	75
HI10	43	L	35	30	45	55	60	70
HI11	18	L	40	35	30	75	70	40
HI12	70	R	45	45	60	55	70	100
Mean	50	8L 4R	34	36	41	46	53	56
St.dev.	17		9	10	9	12	12	22

Table 8.1: *Age, ear measured and pure tone hearing loss of each hearing-impaired subject.*

8.2.2 Stimuli

Non-speech stimuli

The sensitivity to the SAM of a target signal was measured with and without a preceding stimulus to adapt to (adaptor). All stimuli were fully modulated (m=1) or non-modulated (m=0) sinusoids with a carrier frequency of 1 kHz and a fixed duration of 625 ms gated by a cosine-squared window with 5 ms rise/fall times. Where an adaptor was used, the inter-stimulus interval (ISI) was equal to 10, 62, 125, 200, or 625 ms. Each new stimulus was presented at least 1.5 seconds after the previous one and depended on the decision time of the subject. If the adaptor signal was modulated, a sinusoidal amplitude modulation of 8 Hz was applied starting at maximal amplitude (starting-phase of 90 degrees). The signals were presented at 25 dB SL for all subjects. In addition, extra tests were conducted for normal hearing subjects at 50 dB SL to approach equivalent sound pressure levels of the experiments for hearing-impaired subjects (the average hearing loss was 41 dB HL). Stimuli were scaled back to equal RMS.

Speech stimuli

Speech intelligibility was determined using the VU98 sentences spoken by a female speaker (Versfeld et al., 2000). Speech intelligibility was determined in a continuous and in a fluctuating background noise, both having the long-term averaged spectrum of the female speaker. The fluctuating background noise was produced as described by Festen and Plomp (1990). Briefly, speech without pauses is split into two bands (cut-off frequency 1 kHz). The original fine structure is replaced by noise with the same long-term spectrum as speech. This resulted in a noise with a fluctuating behavior similar to speech. Speech was presented at 65 dBA, unless the averaged hearing loss for 500 Hz, 1 kHz and 2 kHz was larger than 40 dB, in which case the speech was presented at 20 dB above the average hearing loss.

8.2.3 Procedure

The following experiments were conducted in a sound-attenuating booth.

Measurement of absolute thresholds

The audibility of a 1 kHz signal was determined using a 1I-2AFC experiment. The interval was indicated using visual cues. Subjects were asked to indicate whether they heard the signal. The signal level was decreased by 4 dB after three successive correct responses and increased by 4 dB after one incorrect response, according to a 1-up 3-down rule, tracking down the 79.1 % point of convergence (Levitt, 1971). After two reversals, the step size was reduced to 2 dB. The threshold was determined by taking the average of the last 6, of 10, reversals. Measurements were carried out twice. When test and retest deviated by more than 3 dB, a third test was carried out. Visual feedback was given for 300 ms after each response.

Sensitivity to the modulation depth of SAM signals

The sensitivity to a SAM signal was determined using a 1I-2AFC procedure. The modulation depth of the target was decreased by 4 dB (20 $\log_{10}[m]$) after three successive correct responses and increased by 4 dB after one incorrect response. After two reversals, the step size was reduced to 2 dB. Hence, the level of 79.1 % correct responses on the psychometric curve is determined. The threshold was determined by taking the average of the last 6, of 10, reversals. Measurements were carried out twice. When test and retest deviated by more than 3 dB a third test was carried out. Visual feedback was given for 300 ms after each response.

Speech intelligibility

Speech intelligibility was measured using a standard speech reception threshold (SRT) test (Plomp and Mimpen, 1979) that determines the speech-to-noise ratio corresponding to 50% correct responses for short meaningful redundant sentences. The sentences (VU98; Versfeld et al., 2000) were presented in a continuous and a fluctuating noise, both with the same long-term averaged noise spectrum as the sentences. When each word in the sentence was (orally) repeated correctly, the speech level was decreased by 2 dB. An incorrectly repeated sentence resulted in a 2 dB higher speech level. The first of 13 sentences was repeated until the subject repeated it correctly, using 4 dB steps in order to quickly converge to 50% intelligibility. Speech intelligibility was estimated by averaging the S/N ratio of the last 9 sentences in noise and the estimated S/N ratio of the sentence following the last sentence. Measurements were carried out twice. When test and retest deviated by more than 2 dB a third test was conducted.

8.2.4 Apparatus

Speech stimuli were reproduced using a compact disc and digitized at a 44.1 kHz sampling rate. The non-speech stimuli were digitally generated at a sampling rate of 16 kHz using a Tucker-Davis Technology (TDT-II) system. Speech and non-speech stimuli were presented via a 16-bit DA-converter (DA3-2) and an anti-aliasing filter (FT6; cut-off frequency 8 kHz). Presentation levels were controlled using programmable attenuators (PA4) and in the case of speech intelligibility, speech and noise were summed using a summer (SM3). These signals passed a headphone buffer (HB6) before being presented to the subjects via headphones (Telephonics TDH 39-P).

8.3 Results

8.3.1 Experiment 1: Sensitivity to SAM

The accuracy obtained in assessing the sensitivity to SAM is high (Cronbach's α = 0.93). According to a paired t-test, there is no significant learning effect (LME, p-value = 0.37).

Group	No adap. (exp. 1)	Unmod. adap. (exp. 2)	Mod. adap. (exp. 2)	Effect
NH(25 dB SL)	-17.2 ± 2.8	-17.5 ± 2.6	-15.9 ± 2.8	1.6 ± 1.8
NH(50 dB SL)	-21.9 ± 3.6	-22.3 ± 3.4	-20.3 ± 4.6	2.0 ± 2.3
HI(25 dB SL)	-21.5 ± 3.3	-20.4 ± 4.4	-15.9 ± 5.5	4.4 ± 3.5

Table 8.2: *Sensitivity to SAM (in dB) for the three groups. The results for experiment 1 are given in the first column, the results of experiment 2 in the last three columns.*

The sensitivity to SAM for normal hearing subjects increased significantly by 4.7 dB (see first column Table 8.2) when the sensation level (SL) is increased from 25 dB SL to 50 dB SL (LME, p-value < 0.01). This is in agreement with Kohlrausch et al. (2000). At equal SL, hearing-impaired subjects are significantly more sensitive to SAM than normal hearing subjects (5.3 dB; p-value < 0.01), whereas when comparisons are carried out at roughly equal SPL-levels, sensitivity is not significantly different (LME, p-value = 0.67).

8.3.2 Experiment 2: Sensitivity to SAM when preceded by an adaptor

The accuracy of the SAM thresholds preceded by an adaptor is high (Cronbach's α = 0.96). According to a paired t-test, there is no significant learning effect (p-value = 0.17).

parameter	NH (25 & 50 dB SL)	NH & HI (equal SL)	NH & HI (comp. SPL)
hearing	n.a.		
SL	4.6**	n.a.	n.a.
ISI			
m1	1.8**	2.2**	3.1**
hearing * ISI	n.a.	*	
hearing * m1	n.a.	**	**
SL * ISI		n.a.	n.a.
SL * m1		n.a.	n.a.
ISI * m1			
hearing * ISI * m1	n.a.		
SL * ISI * m1		n.a.	n.a.

Table 8.3: *Results for the linear mixed effects models The effect of presentation level for normal hearing subjects is given by the first column. The effect of normal hearing and hearing-impaired subjects tested at equal SL and comparable SPL, is given in the second and third column, respectively. The effect-size is given in dB for the main effects. Asterisks denote the level of significance (** p-*value* $<$ *0.05; *** p-*value* $<$ *0.01) The modulation depth of the adaptor is given by m1. The effect of the inter-stimulus interval by ISI. n.a. stands for not applicable.*

Results in Table 8.3 were determined by averaging over subjects and ISIs. The results in Table 8.3 indicate that normal hearing subjects are less sensitive to SAM at 25 dB SL than at 50 dB SL when the target was preceded by an adaptor averaged over ISIs (LME, p-value < 0.0001: 4.6 dB). The sensitivity to SAM is, independent of presentation level, not significantly different for normal hearing and hearing-impaired subjects (LME main effect, p-value $= 0.35$ and p-value $= 0.09$: for 25 SL and at comparable SPLs, respectively). The sensitivity to SAM of the target preceded by a non-modulated adaptor is significantly higher than when the target was preceded by a modulated adaptor (LME, p-value < 0.001). When hearing-impaired and normal hearing subjects were compared using a modulated adaptor and at equal SL, the sensitivity to SAM of the two groups was approximately equal. However, if normal hearing and hearing-impaired subjects are compared at comparable SPLs, the sensitivity to SAM is slightly higher than the sensitivity of hearing-impaired for a non-modulated adaptor. Averages are given in the last three columns of Table 8.2.

Since the results differed greatly between subjects, individual results are given in Figures 8.1, 8.2, and 8.3. The horizontal lines indicate the sensitivity to SAM without an adaptor (Exp. 1). Figure 8.1 gives the results for 10 normal hearing subjects, measured at 25 dB SL. Generally, the sensitivity to SAM preceded by a modulated (plus symbols) or by an non-modulated (circles) adaptor is comparable to the sensitivity to SAM without an adapting stimulus (paired t-test: p-value $>$ 0.1). For relatively short ISIs, the sensitivity to SAM preceded by a non-modulated adaptor is higher than the sensitivity to SAM when preceded by a modulated adaptor (paired t-tests, p-value < 0.01; ISI=10, 62 ms). Three subjects, NH2, NH3 and NH10, deviate from this general trend. For these subjects, particularly NH3, the sensitivity to the SAM of a target is clearly decreased over the whole range of ISIs when a modulated adaptor is present.

Data obtained for normal hearing subjects at 50 dB SL are given in 8.2. The sensitivity to SAM with a non-modulated adaptor (circles) is generally comparable to the sensitivity to SAM without an adaptor (straight line) (paired t-test, p-value = 0.07). In addition, the sensitivity to SAM is often lower when a modulated adaptor (plus symbols) was used instead of a non-modulated adaptor (circles) for all ISIs (paired t-test, p-value < 0.01; except for 125 ms p-value $= 0.45$). NH3 also deviates from the other subjects, there the sensitivity to SAM for NH3 is clearly higher when a non-modulated instead of a modulated adaptor precedes the target.

Figure 8.3 shows that for most hearing-impaired subjects, the sensitivity to SAM without an adaptor is generally similar to SAM when a non-modulated adaptor preceded the target (paired t-tests, p-value > 0.04). The sensitivity to SAM without an adaptor is significantly higher than the sensitivity to SAM when preceded by a modulated adaptor (paired t-tests, p-value < 0.005 for all ISIs). All hearing-impaired subjects, except for HI2, whose results were inconsistent, showed a clearly reduced sensitivity to SAM when the target was preceded by a modulated rather than a non-modulated stimulus. Figure 8.3 also indicates that the individual hearing-impaired subjects react differently to the adaptation stimulus. For instance, HI6 shows an enormous reduction in the sensitivity to SAM when the adaptor was modulated instead of non-modulated, whereas the sensitivity to SAM for other hearing-impaired subjects, such as HI8 and HI5, is barely affected by the modulated adaptor.

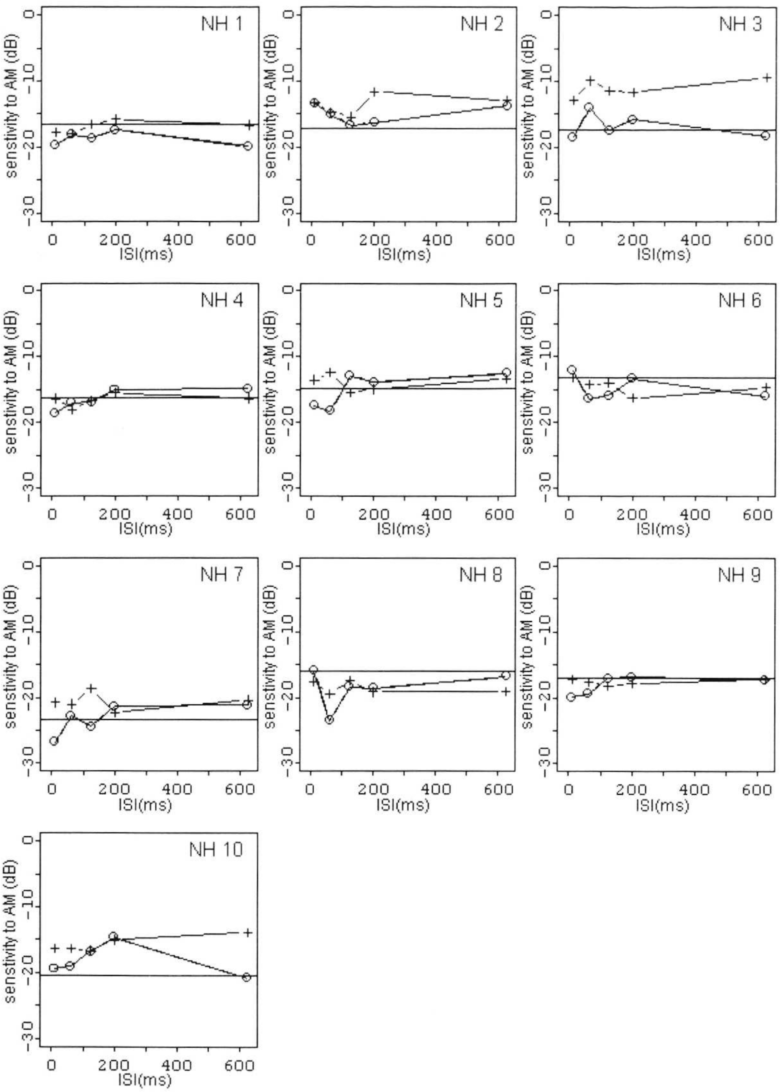


Figure 8.1: *Individual results for the normal hearing subjects measured at 25 dB SL. Sensitivity to SAM (ordinate) is given as a function of the inter stimuli interval in ms (abscissa). The solid horizontal line indicates the sensitivity to SAM without an adaptor (Exp. 1), circles and plus symbols show the sensitivity to SAM when a non-modulated or a modulated adaptor precedes the target, respectively. Subject numbers are given in the right-upper corner.*

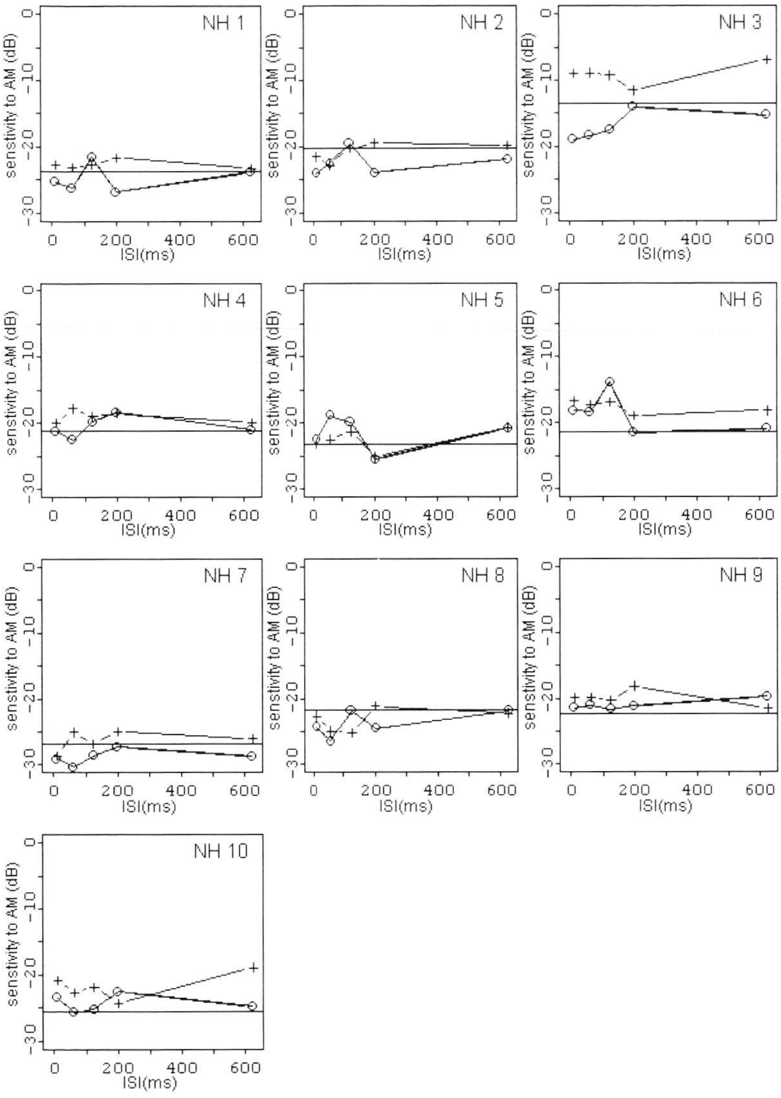


Figure 8.2: *Similar to Figure 8.1 for normal hearing subjects measured at 50 dB SL.*

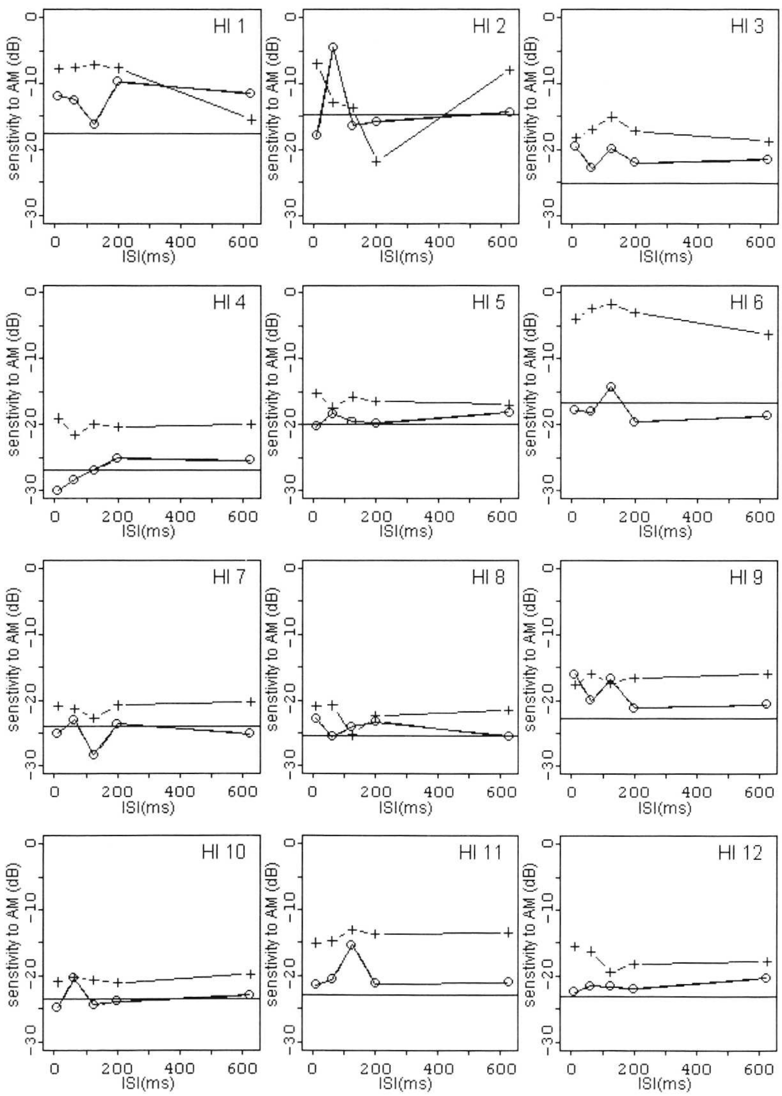


Figure 8.3: *Similar to Figure 8.1 for hearing impaired subjects.*

8.3.3 Experiment 3: Masking release for speech

The accuracy of measurements is very high. Cronbach's α equals 0.98. According to a paired t-test, there were no significant learning effect (p-value = 0.91).

The signal-to-noise ratio at which 50% of the sentences were intelligible is significantly lower for normal hearing subjects than for hearing-impaired subjects (LME, p-value < 0.001; 4.5 dB). The interaction between hearing capacity and the type of noise used is significant (LME, p-value < 0.001). For hearing-impaired subjects, the SRT in a continuous background noise is comparable to the SRT in a fluctuating background noise. For normal hearing subjects, speech intelligibility is significantly better in a fluctuating noise than in a continuous background noise (p-value < 0.001; 4.0 dB). Figure 8.4 shows a scatter plot of the S/N ratios obtained in a continuous background noise (abscissa) and in a fluctuating background noise (ordinate). The horizontal lines indicate the average speech intelligibility in a fluctuating background noise for normal hearing (dashed) and hearing-impaired subjects (solid); the vertical lines indicate the averaged intelligibility of speech in a continuous background noise. The difference in intelligibility between normal and hearing-impaired listeners (distance between the two lines) is clearly larger for the speech presented in a fluctuating noise than in a continuous noise. The sloping line indicates the values where the SRT in fluctuating noise is equal to the SRT in a continuous background noise. The results are comparable to the data for normal hearing and hearing-impaired subjects as reported by Versfeld and Dreschler (2002) indicating that the SRT-values in continuous noise are comparable to the SRT-values in a fluctuating noise unless the SRT in continuous noise is approximately lower than -2 dB. In this case, the SRT in fluctuating noise is better than the SRT in continuous noise.

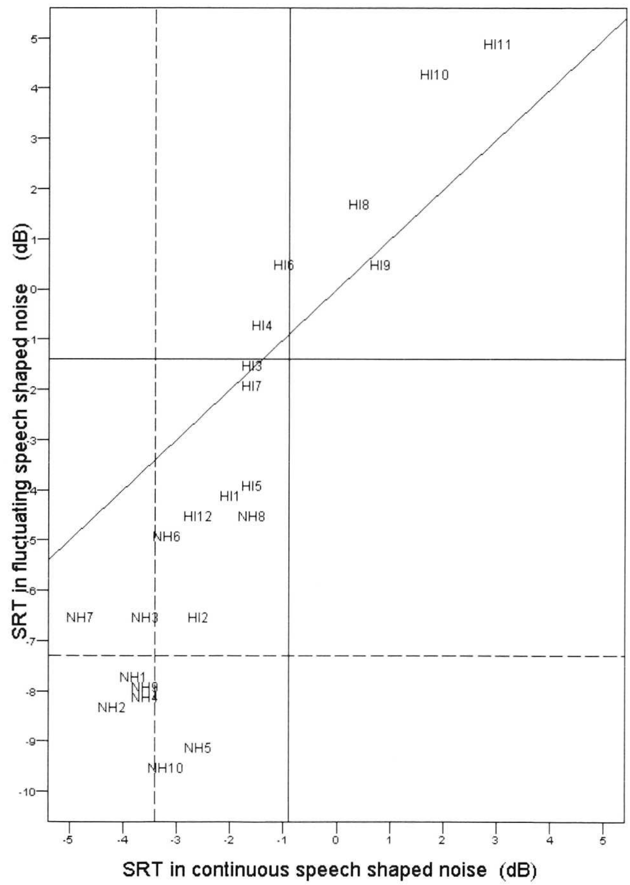


Figure 8.4: *Individual speech intelligibility data. Speech intelligibility obtained in a continuous background noise (abscissa) versus speech intelligibility in a fluctuating background noise (ordinate). The horizontal lines indicate the average S/N ratios for a fluctuating background noise for normal hearing (dashed) and hearing impaired subjects (solid); the vertical lines indicate the average S/N ratios in a continuous background noise.*

8.4 Discussion

8.4.1 Adaptation to SAM

Adaptation to SAM has been observed in psychophysical (Regan and Tansley, 1979; Tansley and Suffield, 1983; Wojtczak and Viemeister, 2003) and physiological (Mäkelä et al., 1987; Coombs and Fay, 1985) experiments for relatively long exposure times to SAM. Using modulation matching after 10 minutes of pre-exposure to SAM (m=0 dB), indicated that the perceived modulation depth for normal hearing subjects was reduced for signals modulated by the same rate using modulation depths of -4, -6, and -8 dB. The amount of adaptation measured by this matching paradigm was similar to the amount reported for modulation detection after pre-exposure to SAM (Wojtczak and Viemeister, 2003). The study described in this chapter shows that subjects, particularly those with impaired hearing, are less sensitive to SAM when it is proceeded by a modulated signal, even when the duration of the adaptor was as short as 625 ms. This suggests that modulation detection interference applies to both simultaneous and non-simultaneous presentation of target and modulated masker.

The results in this chapter indicate that subjects, who are more sensitive to SAM, are also more sensitive to SAM preceded by a non-modulated or another SAM signal. This has been expressed in the correlations between the ability to detect SAM, when it is presented alone and when it is preceded by a non-modulated signal, for both equal SLs ($r = 0.866$, p-value < 0.001) and comparable SPLs ($r = 0.835$, p-value < 0.001). Corresponding correlation coefficients for SAM detection preceded by another SAM signal are 0.581 for equal SLs (p-value < 0.01) and 0.777 for comparable SPLs (p-value < 0.01). There is no association between the adaptation effect, defined as the difference in sensitivity to SAM for non-modulated and SAM-modulated adaptors, and the sensitivity to SAM without an adaptor for equal SLs ($r = 0.266$, p-value > 0.05) or comparable SPLs ($r = -0.281$, p-value > 0.05). The definition of the adaptation effect may result in overestimation for short ISIs, since the non-modulated signal may have served as a reference, improving SAM detection. An 1-AFC procedure was chosen instead of the more commonly used 2- or 3-AFC procedures. This procedure reduces the possibility of comparing the target with other non-modulated signals in the 2- or 3-AFC tasks for the conditions where the non-modulated masker precedes the target. For the group of subjects as a whole, there was no difference between the detection of SAM preceded by a non-modulated signal for ISIs of 10 ms and 625 ms (paired t-test, p-value = 0.07), but some subjects clearly benefited from the non-modulated adaptor for short ISIs, meaning that their reduction in sensitivity to SAM may have been overestimated (*e.g.* HI4, NH1, NH4, and NH5 measured at 25 dB SL and NH1, NH2, NH3, NH7, and NH8 measured at 50 dB SL).

There are three theories in classical psychophysics, which may explain the adaptation effect. Firstly, the human ear may group sounds perceptually. Although modulation detection interference is known to be partly a result of perceptual grouping (Oxenham and Dau, 2001; Bacon et al., 1995; Shailer and Moore, 1993), modulation masking still occurred, reducing the effects known to enhance perceptual grouping, such as rise-fall time differences and similar modulation rates. This implies that, besides the fact that subjects could easily distinguish between the two separate stimuli, perceptual grouping may not have caused these adaptation effects. In addition, when the adaptor is perceived to continue at a modulation rate of 8 Hz, the envelopes of the adaptor and target signals are in phase for ISIs of 125 and 625 ms and out of phase for an ISI of 62 ms. However, there is no significant difference in sensitivity to SAM for ISIs of 62 ms and 125 ms, indicating that the perceptual grouping does not contribute importantly to the adaptation effect.

Secondly, classical forward masking may lead to a reduced sensitivity to SAM, since hearing impaired subjects cannot use the whole signal to detect SAM, as they experienced a slower recovery from forward masking at equal SLs (Glasberg et al., 1987). Since the adaptor and target are presented at the same SL and the ISI are longer than the periods in which forward masking may occur (Duifhuis, 1973), it is unlikely that forward masking causes the adaptation effect.

Thirdly, a loss of compression enhances a modulated envelope, meaning that the adaptation effect may be larger for hearing-impaired subjects. However, since adaptation was defined in terms of the difference between the sensitivity to SAM preceded by a non-modulated and by a modulated adaptor, the target will be subjected to the same compression system in both cases. In addition, a signal cannot be more than fully modulated, which implies that the induced sensation is probably similar.

The results also indicate that the recovery period may last longer than previously thought. The reduction in sensitivity caused by a modulated adaptor with ISIs of 10, 62 and 625 ms is not significantly different for the whole group of subjects (paired t-tests, p-values > 0.1). However, the subgroup of hearing-impaired subjects with a large adaptation effect (>3 dB) became more sensitive to SAM for increasing ISI. Exposure times of 20 minutes may require recovery periods of almost 1 minute (Tansley and Suffield, 1983), suggesting that adaptation effects observed in periods of up to 625 ms are not particularly unusual.

The results presented in this chapter indicate that hearing-impaired subjects need longer to process the modulation of an adaptor. They were presented with a modulated target, while their auditory system was still analyzing the first signal. They are only able to detect AM if the modulation depth is larger than that of the remainder of the 'after image' of the adaptor. Physiologically, these results can be interpreted as indicating that hearing-impaired subjects' SAM-filters (as suggested by Millman et al., 2002) recover more slowly, adapt more quickly, or do both. Psychophysically, the analysis window of the hearing impaired ear may be inadequate.

8.4.2 The relationship between release of masking for speech and the sensitivity to SAM

The sensitivity to SAM is a measure of temporal resolution and determines the accuracy, by which the envelope of a signal can be detected. Since the envelope is very important to the intelligibility of speech, the sensitivity to SAM may correlate with speech intelligibility. However, there seems to be no relationship between speech intelligibility in fluctuating noise and SAM detection without an adaptor at equal SLs (r = 0.22, p-value = 0.5) or at comparable SPLs (r = 0.152, p-value > 0.32). The same holds for SAM measurements preceded by an non-modulated or a modulated adaptor (r < 0.20; p-value > 0.45). Speech intelligibility in a continuous background noise correlates significantly with the sensitivity to SAM without an adaptor measured at equal sensation levels (r = -0.48; p-value = 0.024). However, the correlation disappears when the SII (ANSI, 1997), instead of the actual SRT, is used. Since the SII incorporates the effect of hearing loss in intelligibility scores, this indicates that the use of two non-homogeneous subgroups, consisting of normal hearing and hearing-impaired subjects, is the most important factor in the significance of the correlation. In addition, the SRT in continuous noise and the SII do not correlate with the sensitivity to SAM preceded by an adaptor ($r<-0.348$; p-value > 0.11).

Experiment 2 shows that placing a modulated adaptor before the target reduced the sensitivity to SAM, especially for hearing-impaired subjects. Experiment 3 clearly indicates that normal hearing subjects experience better speech intelligibility in the fluctuating noise than in the continuous noise, whereas this is often not true for hearing-impaired subjects (the first two rows of Table 8.4). Figure 8.5 illustrates the results for all subjects for both experiments. The amount of adaptation (abscissa) is largest for hearing-impaired subjects, especially HI6. The amount of masking release for speech (ordinate) is smaller for hearing-impaired subjects, especially HI6 and HI10. Apparently, the masking release for speech reduces when the size of the adaptation effect, at equal SLs averaged over all ISIs (abscissa), increases (*i.e.* more masking).

In the previous section, it was mentioned that the adaptation effect may have been overestimated since the non-modulated adaptor may have served as a reference. The third and fourth row of Table 8.4 represent the correlation coefficients between the release of masking for speech and the size of the adaptation effect, defined as the difference between the sensitivity to SAM without an adaptor and the sensitivity to SAM preceded by a modulated adaptor. The relationship between this definition of the adaptation effect averaged over all ISIs and the release of masking for speech is illustrated in Figure 8.6. The proportion of variance explained was 0.36 for the original definition of the adaptation effect and 0.50 for the revised definition. The last two rows in Table 8.4 indicate that the correlations based on hearing-impaired subjects alone are reasonably high, implying that the relationships found are not the results only due to the comparison between normal hearing and hearing-impaired subjects.

A lot of studies have examined the relationship between temporal resolution and speech intelligibility. Most studies indicate that speech intelligibility is only marginally dependent on measures of temporal resolution. In an extensive study by Festen and Plomp (1983), significant correlation coefficients were reported between

	ADAPTATION					
	ISI_{10ms}	ISI_{62ms}	ISI_{125ms}	ISI_{200ms}	ISI_{625ms}	$\overline{ISI}$
MR (SL) (N=22)	-0.27	-0.51*	-0.48*	-0.60**	-0.42	-0.60**
MR (SPL) (N=22)	-0.34	-0.57**	-0.34	-0.59**	-0.45*	-0.59**
MR (SL re. dlm) (N=22)	-0.63**	-0.69**	-0.64**	-0.76**	-0.62**	-0.70**
MR (SPL re. dlm) (N=22)	-0.55**	-0.63**	-0.52*	-0.68**	-0.52*	-0.65**
MR (HI) (N=12)	-0.01	-0.75**	-0.09	-0.67*	-0.51	-0.57
MR (HI) (re. dlm)(N=12)	-0.21	-0.50	-0.43	-0.65*	-0.51	-0.64*

Table 8.4: *Correlation coefficients between the amount of adaptation and the release of masking for speech. The columns give the ISI at which adaptation is determined. SL and SPL indicates that the normal hearing data were obtained at 25 dB SL and 50 dB SL, respectively. Asterisks denote the level of significance (** p-*value* < *0.05;* ** p-*value* < *0.01)*

speech intelligibility in noise and spectral measures, but not for temporal measures such as forward masking, backward masking, click thresholds, and click thresholds in noise. This indicates that, spectral properties are more closely related to speech intelligibility in continuous background noise. Takahashi and Bacon (1992) found no significant correlations between modulation detection and the intelligibility of speech in noise. Dreschler and Leeuw (1990) found that, from a selection of conditions, only the STI and the minimal detectable gap thresholds for wideband signals correlated significantly. The significant correlations found in the present study suggest that the extent to which subjects can benefit from temporal gaps in the noise is probably more related to measures of temporal resolution, at least for adaptation to SAM, than speech intelligibility in a continuous background noise. When speech is presented in a fluctuating background noise, similar effects may occur as when SAM is preceded by a modulated adaptor. The results suggest that hearing-impaired subjects need more time to process the information in an adaptor or a fluctuation in the background

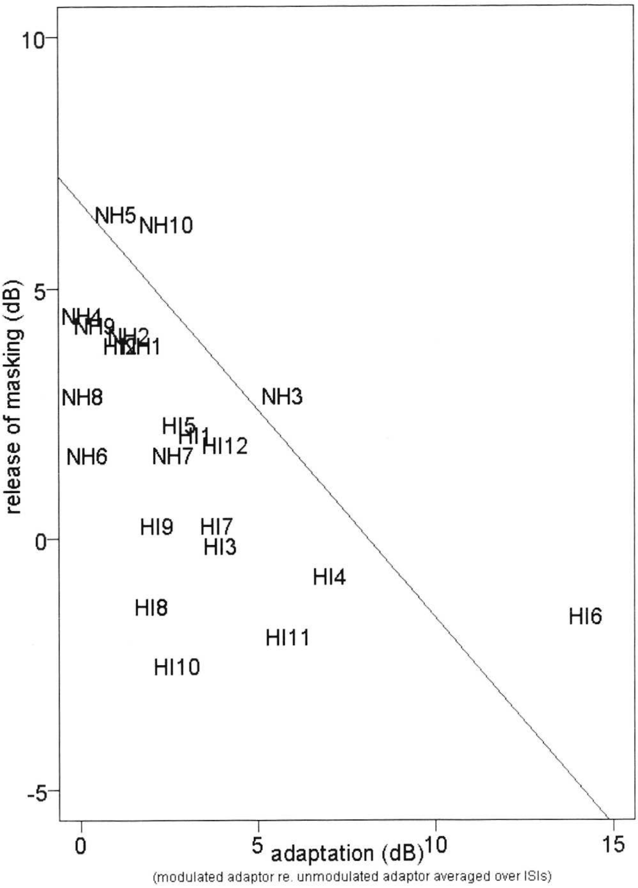


Figure 8.5: *Scatter plot of the release of masking and the adaptation effect. The difference in sensitivity to SAM preceded by a modulated adaptor and a continuous adaptor is given on the abscissa (averaged over all ISIs). The release of masking for speech in fluctuating noise compared to continuous noise is given on the ordinate.*

noise. Similar problems may occur for normal hearing subjects for simultaneous MDI. This may explain why normal hearing and hearing-impaired subjects encounter the same amount of interference (Grose and Hall, 1994). In order to improve SAM detection and speech intelligibility, the modulation depth would have to be increased or the level of noise reduced. This would allow the signal to be processed adequately.

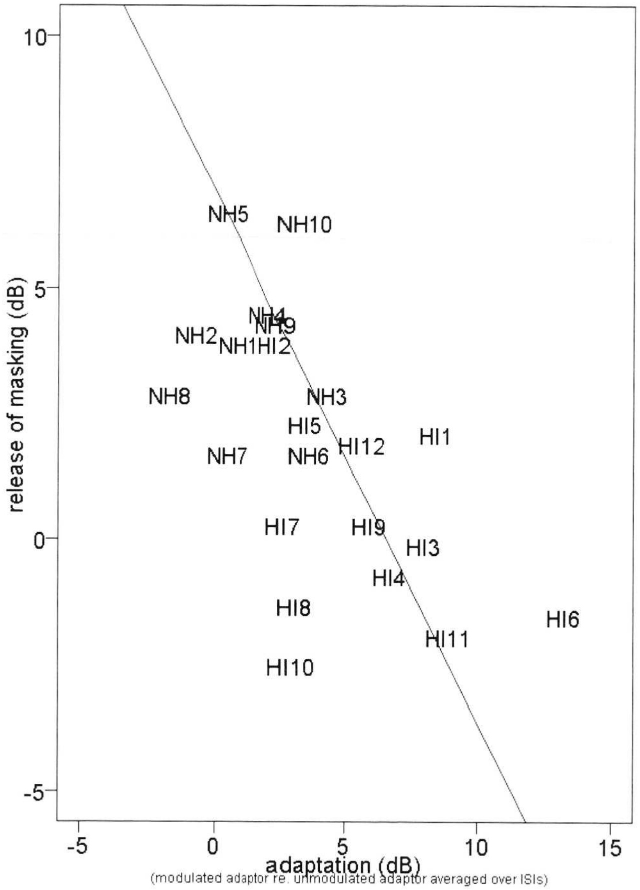


Figure 8.6: *As in Figure 8.5, but now the amount of adaptation is determined relative to the situation without a preceding stimulus*

8.5 Conclusions

This study has examined whether the sensitivity to SAM is reduced when a SAM signal is preceded by a modulated signal. The results indicate that:

1. modulation detection interference (MDI) occurs in signals presented simultaneously as well as those presented non-simultaneously.

2. non-simultaneous MDI, or adaptation, appears to occur for all hearing-impaired and some normal hearing subjects.

3. the amount of adaptation correlates strongly with the release of masking for speech. Apparently, the hearing-impaired ear needs longer to recover from pre-exposure to SAM.

4. adaptation of modulation perception may be one of the factors, which prevent hearing-impaired listeners from profiting from the temporal gaps in background noise.

Chapter 9

From simultaneous to non-simultaneous MDI for normal hearing and hearing-impaired subjects*

Abstract

The sensitivity to SAM is reduced when other modulated maskers are presented simultaneously at a distant frequency (MDI). In Chapter 8, similar interference is reported for non-simultaneous presentation, especially for hearing-impaired subjects. This is referred to as adaptation. This chapter investigates the interference for normal hearing and hearing-impaired subjects for conditions that gradually change from simultaneous to non-simultaneous presentation. Subjects were asked to detect the SAM of a 625 ms 1 kHz carrier modulated by 8 Hz sinusoid. The presentation level was 25 dB SL. For some normal hearing subjects, measurements were also performed at 50 dB SL. The target was presented alone or in the presence of a 2 kHz masker that was either non-modulated or modulated by the same modulation rate (in

*submitted to the Journal of Acoustical Society of America

phase). The masker was presented simultaneously, partly simultaneously, and non-simultaneously. Large differences in the reduction in sensitivity to SAM were found for normal hearing and hearing-impaired subjects for simultaneous presentation of target and masker. For both, normal hearing and hearing-impaired subjects, an asymptotically reduced sensitivity to SAM was found as a function of ISI. Significant correlation coefficients were found between MDI and the benefit to the intelligibility of speech presented in a fluctuating noise rather than a continuous noise.

9.1 Motivation

Modulation perception in the presence of modulated maskers has been determined for two situations:

1. Simultaneous presentation of target and masker using different carrier frequencies (MDI: Chapter 5)
2. Non-simultaneous presentation of target and masker using similar carrier frequencies (adaptation: Chapter 8)

For MDI, the performance of hearing-impaired subjects was similar to that of normal hearing subjects, whereas for adaptation large differences were reported between normal hearing and hearing-impaired subjects. The exact nature of these differences is unclear. This study was designed to investigate the effect of simultaneous and non-simultaneous presentation of modulated maskers on the sensitivity to SAM.

9.2 Methods

9.2.1 Subjects

Eight normal hearing and eight sensorineurally hearing-impaired subjects participated in this study. The normal hearing subjects had thresholds better than 15 dB HL at the standard audiometric frequencies from 250 Hz to 8 kHz (re. ANSI, 1996). The average age was 29 years and normal hearing subjects were tested at a randomly assigned ear. The age, ear tested and auditory thresholds of the hearing-impaired subjects are presented individually in Table 9.1. The bottom row of Table

9.1 presents the average age and thresholds of the hearing-impaired subjects, with their standard deviations. The subject numbers were allocated according to an increasing average loss at the frequencies 1, 2, and 4 kHz. In case of an asymmetrical hearing loss, the best ear was tested.

			Pure tone thresholds					
Subject	Age	Ear	0.25 kHz	0.5 kHz	1 kHz	2 kHz	4 kHz	8 kHz
HI1	59	L	30	35	40	40	40	55
HI2	64	R	45	45	40	30	60	75
HI3	64	R	30	30	35	40	60	65
HI4	62	L	30	45	50	50	50	40
HI5	62	L	30	40	45	50	60	65
HI6	59	L	15	10	45	60	55	75
HI7	18	L	40	35	30	75	70	40
HI8	70	R	45	45	60	55	70	100
Mean	57	5L 3R	33	36	43	50	58	64
St. dev.	16		10	12	9	14	10	20

Table 9.1: *Age, ear tested and pure tone audiogram for the hearing-impaired subjects. The last row shows the averages and standard deviation between brackets.*

9.2.2 Stimuli

Non-speech stimuli

The sensitivity to SAM in a target signal was measured for three conditions:

1. the sensitivity to SAM in an 8 Hz modulated, 1 kHz pure tone carrier (PA).
2. the sensitivity to SAM of an 8 Hz modulated 1 kHz pure tone presented in the presence of a non-modulated 2 kHz masker (UNMOD).
3. similar to 2, but the 2 kHz masker was 100% sinusoidally amplitude modulated using a modulation rate of 8 Hz (MOD).

All stimuli had a fixed duration of 625 ms, gated by a cosine-squared window with 5 ms rise/fall times presented at equal RMS-levels. When a masker was present, the inter-stimulus interval (ISI) varied for different experiments and was completely simultaneous (ISI = -625 ms), partially simultaneous (ISI = -500, -375, -125 ms), or completely non-simultaneous (ISI = 0 and 125 ms). The envelope of the target and masker always started at the DC-component and were in phase†. The time between two stimuli depended on the subjects' decision time, but was at least 1.5 seconds. The signals were presented at 25 dB SL for all subjects. For six normal hearing subjects additional tests were conducted at 50 dB SL to approach equivalent sound pressure levels to those used for the hearing-impaired subjects, who had an average hearing loss of 43 dB HL.

The stimuli properties given above are the standard values. In Experiment 3, four of these parameters were varied in order to study the effect of perceptual grouping. Firstly, the phase relation between the envelope of the target and the masker was varied. For an ISI of -375 ms, the envelopes were presented either without a phase difference between target and masker (standard) or 180^0 out of phase. Secondly, the center frequency of the carrier was varied. At an ISI of 125 ms, masker carriers used were identical to the target (1 kHz), or as in the standard situation (2 kHz). Thirdly, the modulation depth of the adaptor was varied. The carrier frequency of the target and masker were 1 kHz and the modulation depth of the adaptor was 0, 0.5, or 1. Finally, the modulation rate was varied. The masker was either non-modulated (0 Hz) or fully modulated using a modulation frequency of 4, 8, or 16 Hz. The carrier frequency for the experiments in which the modulation depth or modulation rate of the adaptor was varied, was 1 kHz for target and masker.

Speech stimuli

Speech intelligibility was determined using the VU98 sentences spoken by a female speaker (Versfeld et al., 2000). Speech intelligibility was determined in a continuous and in a fluctuating background noise, both having the long-term averaged spectrum as the female speaker. The fluctuating background noise was produced as described

†Some of the data presented in section 9.3.3 (carrier frequency of the masker at 1 kHz, with modulation depths of 0 and 1) have been reported in Chapter 8. Signal generation for the two experiments differed to a small extent. This experiment used modulators that started at the DC-component (phase = 0^0), whereas the modulators from the previous experiment started at maximum amplitude (phase = 90^0). It is hypothesized that this difference will not influence the data.

by Festen and Plomp (1990). Speech was presented at 65 dBA, unless the averaged hearing loss for 500 Hz, 1 kHz and 2 kHz was larger than 40 dB, in which case the speech was presented 20 dB above the average hearing loss.

9.2.3 Procedure

In a sound-attenuating booth the following experiments were conducted:

Measurement of absolute thresholds

The audibility of a 1 kHz signal was determined using a 1I-2AFC experiment. The interval was indicated using visual cues. Subjects were asked to indicate whether they heard the signal. The signal level was decreased by 4 dB after three successive corrects responses and increased by 4 dB after one incorrect response. After two reversals, the step size was reduced to 2 dB. Hence, the point of 79.1% correct responses was determined (Levitt, 1971). The threshold was determined by taking the average of the last 6, of 10, reversals. Measurements were carried out twice. Visual feedback was given for 300 ms after each response. If test and retest deviated by more than 3 dB, a third test was carried out.

Sensitivity to SAM signals

The sensitivity to SAM of a 1 kHz carrier was determined using the adaptive procedure described above with a initial step size of 4 dB, which was reduced to 2 dB in modulation depth ($20 \log_{10}[m]$) after two reversals.

Speech intelligibility

Speech intelligibility was measured using a standard speech reception threshold (SRT) test (Plomp and Mimpen, 1979), similar to that in Chapter 8, which determines the speech-to-noise ratio corresponding to 50% correct responses for short meaningful redundant sentences (VU98; Versfeld et al., 2000). Sentences were presented in a continuous or a fluctuating noise, both with the same long-term averaged spectrum as the sentences. If each word in the sentence was repeated correctly, the speech level was decreased by 2 dB. If the sentence was not repeated correctly, the speech level was increased by 2 dB. The first of 13 sentences was repeated until the subject repeated it correctly, using 4 dB steps in order to quickly converge to an overall intelligibility of 50%. Speech intelligibility was estimated by averaging the S/N ratio of the last 9 sentences in noise and the estimated S/N ratio of the sentence following the last sentence. Measurements were carried out twice. When test and retest deviated by more than 2 dB a third test was conducted.

9.2.4 Apparatus

Speech stimuli were reproduced from a compact disc and digitized at a 44.1 kHz sampling rate. The non-speech stimuli were generated digitally using a sampling rate of 16 kHz and a Tucker-Davis Technology (TDT-II) system. Speech and non-speech stimuli passed via a 16-bit DA-converter (DA3-2) and an anti-aliasing filter (FT6; cut-off frequency 8 kHz). Presentation levels were controlled using programmable attenuators (PA4) and for speech intelligibility, speech and noise were summed using a summer (SM3). These signals passed a headphone buffer (HB6) before being presented to the subjects via headphones (Telephonics TDH 39-P).

9.3 Results

9.3.1 Experiment 1: Sensitivity to SAM

The accuracy of measurements determining the sensitivity to SAM was very high (Cronbach's $\alpha = 0.93$). There is no significant learning effect (paired t-test, p-value > 0.1).

NH (25 dB SL)	NH (50 dB SL)	HI (25 dB SL)
-17.2 (3.1)	-22.8 (2.1)	-23.5 (2.2)

Table 9.2: *The sensitivity to SAM (dB) determined in Experiment 1*

The direct results are given in Table 9.2. The sensitivity to SAM for normal hearing subjects increased significantly by 5.5 dB when the sensation level (SL) was increased from 25 dB SL to 50 dB SL (LME, p-value < 0.001). At equal SL, hearing-impaired subjects were significantly more sensitive to SAM than normal hearing subjects

(LME, 6.2 dB; p-value < 0.001). There was no significant difference when these groups are compared at roughly equal SPL-levels (LME, p-value=0.49).

9.3.2 Experiment 2: Sensitivity to SAM in the presence of a masker

The accuracy of the SAM thresholds in the presence of a masking stimulus, is very high (Cronbach's $\alpha = 0.96$). There was no significant learning effect (paired t-test; p-value $= 0.29$).

Since the inter-individual differences are quite large, individual results are given for normal hearing subjects measured at 25 dB SL (Figure 9.1), 50 dB SL (Figure 9.2), and for hearing-impaired subjects (Figure 9.3). In these figures, the sensitivity to SAM (ordinate) is given as a function of the inter-stimulus interval (ISI) in ms (abscissa). The horizontal lines give the thresholds for the sensitivity to SAM determined in Experiment 1. Circles and plus symbols represent the sensitivity to SAM presented in a non-modulated and a modulated masker, respectively. Hence, the classical MDI is given by the difference between the plus symbols and circles for simultaneously presented signals (-625 ms). Some subjects experienced difficulties identifying whether the stimulus was modulated for conditions where target and masker are presented simultaneously (ISI = -625 ms). These subjects, five hearing-impaired subjects (HI2, HI4, HI5, HI7, and HI8) and 1 normal hearing subject (NH6), were unable to complete this task, even after four hours of practice. These thresholds are set to 0 for further analysis. In addition, the sensitivity to SAM for subject HI4 was not measured for a non-modulated masker presented simultaneously with the target, since this subject was unable to complete the task in the presence of a modulated masker.

Table 9.3 shows the results of the linear mixed effects models for the parameters in this experiment. The sensitivity to SAM decreased significantly as the ISIs decreased (LME, p-value < 0.001), although NH1 demonstrated barely any interference. This decreased sensitivity to SAM is particularly evident for a modulated masker (plus symbols in Figure 9.1, 9.2 and 9.3). However, the sensitivity to SAM was also lower for non-modulated maskers (circles) than without a masker (Experiment 1) for hearing-impaired subjects (t-test; p-value $= 0.006$), whereas

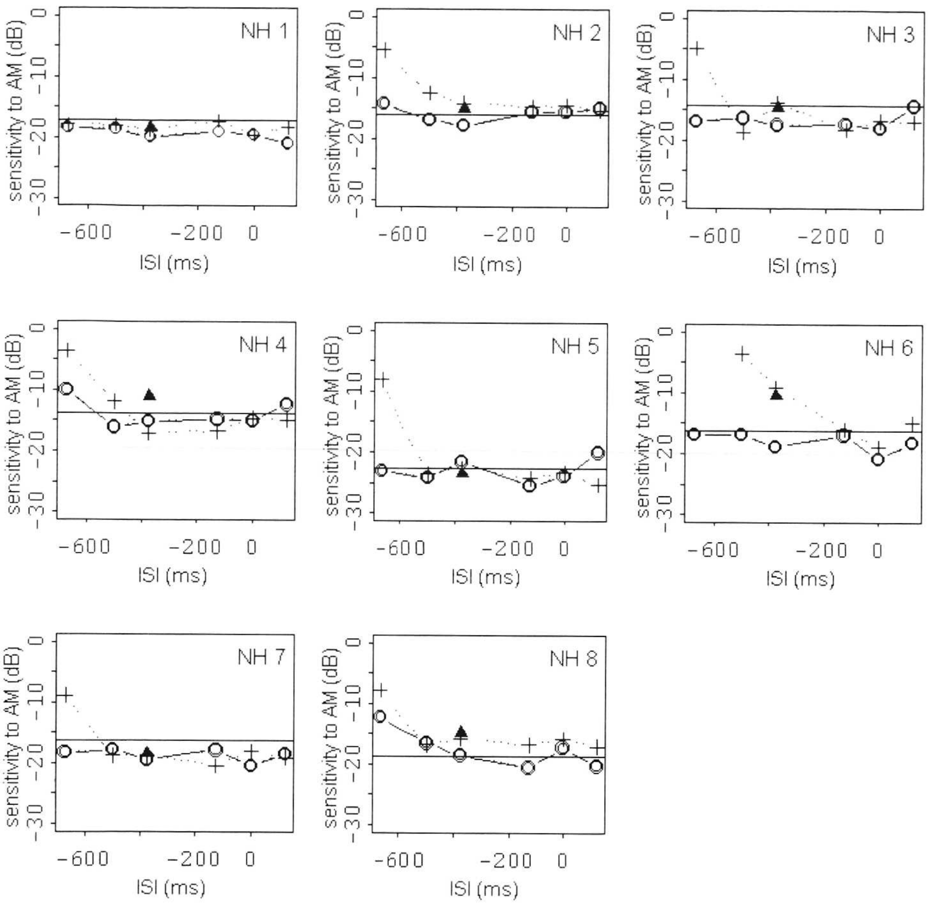


Figure 9.1: *Sensitivity to SAM (ordinate) as a function of the ISI between masker and target (abscissa) for normal hearing subjects measured at 25 dB SL presented together with a modulated (plus symbols) or non-modulated (circles) masker. The horizontal line shows the sensitivity to SAM without added components. Triangles represent the sensitivity to SAM in the presence of a modulated masker with envelopes out of phase (Experiment 3).*

results for normal hearing subjects were similar in both conditions (t-test; p-value > 0.1). The interference effect for hearing-impaired subjects using non-modulated maskers at other ISIs than -625 ms was not significantly different from 0. However, normal hearing subjects measured at 25 dB SL, showed a significant interference for

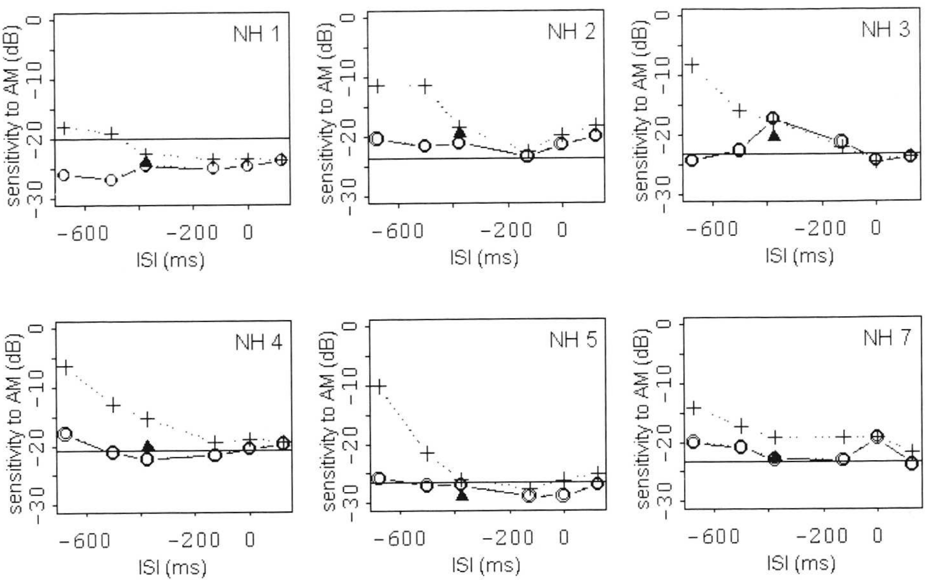


Figure 9.2: *Similar to Figure 9.1, for normal hearing subjects measured at 50 dB SL.*

ISIs of -375, -125 and 0 ms (p-value = 0.02 / 0.005 / 0.037, respectively), indicating that sensitivity to SAM increases in the presence of a non-modulated masker. The significance of sensation level reflects the increased sensitivity of normal hearing subjects to SAM for an increased sensation level.

Figure 9.4 shows the MDI, as determined by the difference in sensitivity to SAM in the presence of a modulated masker and a non-modulated masker (parameter m1) as a function of the time delay between the masker and target (parameter ISI). Note that both signals, target and masker, have an overall duration of 625 ms. Hence, signal and masker are only presented completely simultaneously reflecting classical MDI (Yost and Sheft, 1989), for ISI = -625 ms. For other ISIs (< 0), the signal was partly simultaneous and partly non-simultaneous. For normal hearing subjects measured at 25 dB SL a clear effect of MDI was found for ISI -625 ms (t-test, p-value < 0.001), but not for other ISIs except at 0 ms (t-test, p-value $= 0.004$). Normal hearing subjects measured at 50 dB SL also showed a clear effect of MDI for

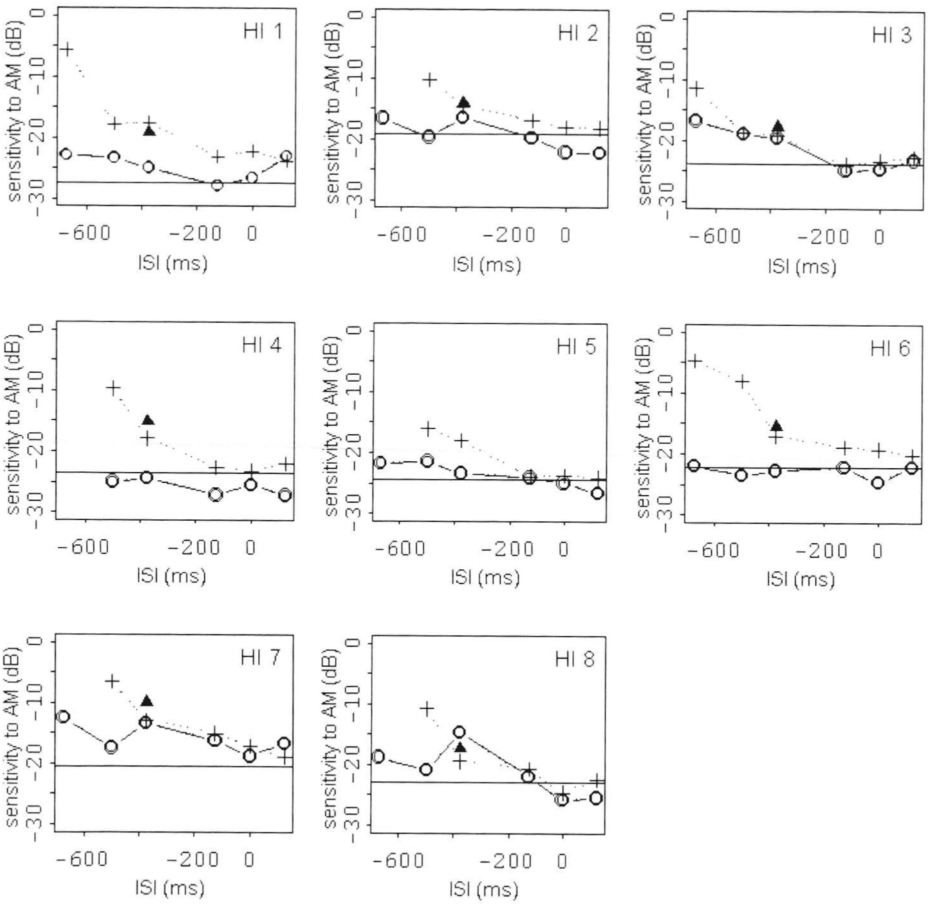


Figure 9.3: *Similar to Figure 9.1, for hearing-impaired subjects, measured at 25 dB SL.*

simultaneous presentation, -625 ms (t-test, p-value < 0.001). However a significant interference effect was also found at an ISI of -500 ms (t-test, p-value < 0.001) but not for other ISIs. For hearing-impaired subjects a significant effect of interference was found for ISIs ranging from -625 ms to 0 ms (t-test, p-value < 0.004), except for ISI = -375 ms (p-value = 0.08). MDI was significantly higher for hearing-impaired subjects than for normal hearing subjects at -625 ms measured at 25 and 50 dB SL (t-test, p-value = 0.003 and p-value = 0.008, respectively). For an ISI of -500 ms

Condition	NH (SL) & HI	NH (SPL) & HI	NH (SL) & NH(SPL)
NH/HI	2.6	1.7	n.a.
ISI	7.0**	8.7**	5.7**
m1	3.2**	3.5**	3.0**
SL	n.a.	n.a.	3.8**
NH/HI * m1	*	0	n.a.
NH/HI * ISI	**	*	n.a.
SL * m1	n.a.	n.a.	*
ISI * m1	**	**	**
SL * ISI	n.a.	n.a.	0
NH/HI * ISI * m1	0	0	n.a.
SL * ISI * m1	n.a.	n.a.	0

Table 9.3: *Interaction table for the results of Experiment 2. For the main parameters the difference in dB modulation depth is given. The first column shows the results when hearing-impaired and normal hearing subjects were compared at 25 dB SL. The second column shows the results when hearing-impaired and normal hearing subjects were compared at comparable SPLs (50 dB SL). The third column shows the results when normal hearing subjects were compared at 25 and 50 dB SL. Asterisks denote the level of significance; 0* p-value *>0.01; ** p-value*<0.01; *** p-value*<0.001. Parameters NH/HI, ISI, m1 and SL stand for hearing capacity, inter-stimuli-interval, masker modulation depth and sensation level, respectively. Conditions that are not applicable are denoted by n.a.*

there was a significantly lower interference for normal hearing subjects measured at 25 dB SL (p-value = 0.005) but not for the data measured at 50 dB SL (p-value = 0.14). The interference for normal hearing subjects at an ISI of -625 ms was not significantly different from the interference at -500 ms for hearing-impaired subjects (p-value = 0.8).

9.3.3 Experiment 3: Additional factors influencing perceptual grouping

As mentioned in section 4.5, it has been suggested that perceptual grouping is one of the principles underlying MDI. In experiment 3, several parameters were studied, which are known to stimulate perceptual grouping. In each subplot of Figure 9.5, a different parameter was studied which promotes either perceptual grouping or

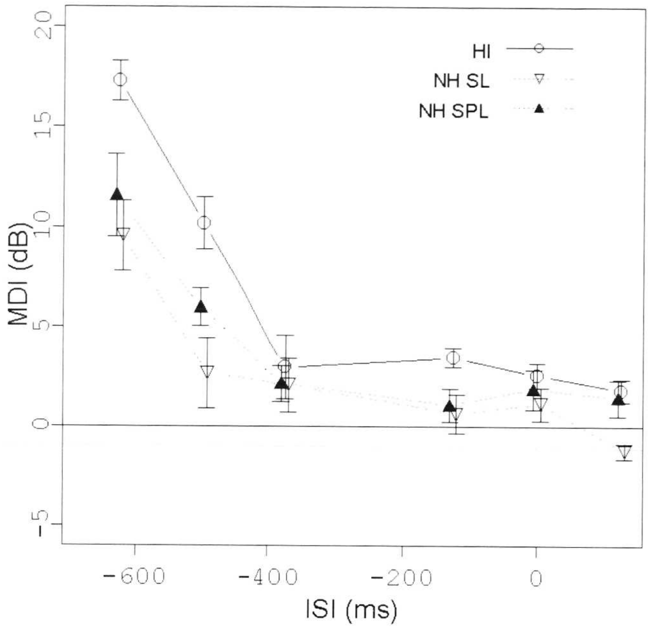


Figure 9.4: *MDI (ordinate) as a function of ISI between masker and target (abscissa) for hearing-impaired subjects (circles), normal hearing subjects measured at 25 dB SL (open triangles) and normal hearing subjects measured at 50 dB SL (solid triangles).*

segregation. Thresholds are expressed relative to the sensitivity to SAM without a masker.

Figure 9.5A shows the influence of a common modulator phase for the simultaneous trajectory of target and masker. For an ISI of -375 ms, the interference was determined for conditions with the envelopes of the target and masker modulated in phase and in anti-phase (counter phase; presented by triangles in Figures 9.1, 9.2, and 9.3). HI5 did not participate in the tests due to time constraints. Diagonally hatched bars represent the condition with non-modulated maskers; vertically hatched bars represent the in-phase condition and horizontally hatched bars represent the condition for the envelopes 180^0 out of phase. There is a clear interference effect caused by each of the three maskers for hearing-impaired subjects and the reduction in sensitivity to SAM is significantly higher when using modulated rather than non-modulated maskers (paired t-test, p-value < 0.001). Thresholds were not significantly lower when the envelope of the masker was modulated out

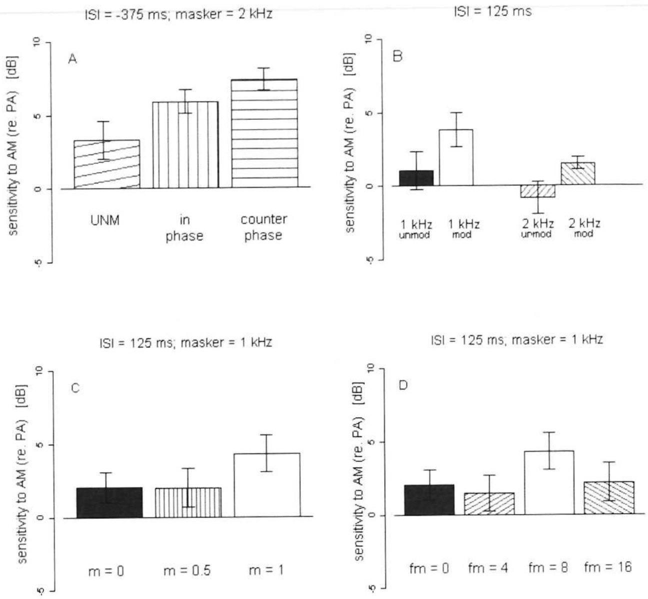


Figure 9.5: *Four parameters influencing perceptual grouping for hearing-impaired subjects.*
A; the influence of a phase difference between the envelope of masker and target.
B; the influence of carrier frequency
C; the influence of the modulation depth of the preceding stimulus.
D; shows the influence of the modulation rate of the preceding stimulus.

of phase instead of in phase with the target (p-value $= 0.04$). There was even a tendency for more interference when the envelopes were modulated out of phase. For normal hearing subjects, there was no significant interference effect using modulated maskers at ISI -375 ms.

Figure 9.5B shows the influence of carrier frequency. The target was presented 125 ms after the masker ended, which was either equal in carrier frequency (1 kHz, open and solid bars) or higher in carrier frequency (2 kHz, hatched bars) than the target. There is no significant interference effect for each of the four conditions for normal hearing subjects (p-value > 0.11). Significant interference was found for hearing-impaired subjects when a modulated masker preceded the target (t-test, p-value < 0.017). There was a tendency for more interference when the masker frequency is equal to the target frequency (p-value $= 0.049$).

Figure 9.5C shows the effect of modulation depth of the masker for an ISI of 125 ms with a masker carrier frequency of 1 kHz, which was modulated by a modulation depth of 0, 0.5 or 1, given by solid, hatched and open bars, respectively. For hearing-impaired subjects there was a significant effect of interference for a fully modulated masker, whereas a modulation depth of 50% did not affect the sensitivity to SAM. The interference effect for normal hearing subjects was not significantly different from 0.

Figure 9.5D shows the influence of common modulation frequency. The target was presented 125 ms after a 1 kHz masker and modulated by either 0 Hz (continuous tone), 4, 8, or 16 Hz, presented by solid, hatched, open and hatched bars, respectively. For hearing-impaired subjects, there is no significant interference effect, except for the condition in which the masker modulation was equal to the target modulation frequency of 8 Hz. For normal hearing subjects, there was no significant interference effect for each of the modulation frequencies. Wojtczak and Viemeister (2003) showed this kind of tuning for the adaptation effect for normal hearing subjects after 10 minutes of pre-exposure to a fully modulated stimulus using a modulation frequency of 8, 12, 16, or 20 Hz for modulation rates for a target modulation rate of 16 Hz.

9.3.4 Experiment 4: Speech intelligibility in noise

The accuracy of the measurements, as measured by Cronbach's α, is high (0.93). There is no significant effect of learning (paired t-test, p-value = 0.13).

Figure 9.6 shows the signal-to-noise ratio at which 50% of the sentences were intelligible. These S/N ratios are significantly lower for normal hearing subjects than for hearing-impaired subjects (LME, p-value < 0.001). The main effect, noise type, is not significant but the interaction between hearing capacity and the noise type is significant (LME, p-value < 0.001). For hearing-impaired subjects, the SRT in a continuous background noise was generally comparable to the SRT in a fluctuating background noise and datapoints are situated around the diagonal line given in Figure 9.6. For normal hearing subjects, speech intelligibility is significantly better in a fluctuating noise than in a continuous background noise (LME, p-value < 0.001; 4.0 dB). The data are comparable to those reported for normal hearing and hearing-impaired subjects by Versfeld and Dreschler (2002). They indicated that the SRT-

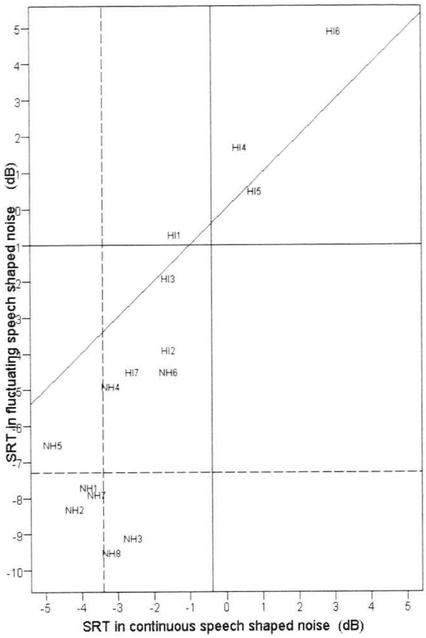


Figure 9.6: *Individual speech intelligibility data. Speech intelligibility obtained in a continuous background noise (abscissa) versus speech intelligibility in a fluctuating background noise (ordinate). The horizontal lines indicate the average S/N ratios for a fluctuating background noise for normal hearing (dashed) and hearing-impaired subjects (solid); the vertical lines indicate the average S/N ratios in a continuous background noise. The sloping line indicates points at which the SRT in a continuous background noise equalled the SRT in a fluctuating background noise.*

values in a continuous background noise were comparable to the SRT-values in a fluctuating noise as long as the SRT in continuous noise was larger than -2 dB. For SRTs in continuous noise below -2 dB, the SRT in fluctuating noise becomes better.

9.4 Discussion

9.4.1 Masking of SAM

Hall and Grose (1991) conducted MDI experiments at which the masker and target were gated on simultaneously and non-simultaneously. The carriers and the exposure level (70 dB SPL) were comparable to those used in this study at 50 dB SL. For simultaneous gating. MDI was approximately 9 dB, similar to this study (ISI = -625 ms: 11 dB standard error 1 dB). For asynchronous gating, starting after 1 cycle of the envelope, they still found an interference of 2-3 dB similar to the effect for a similar condition (ISI = -500 ms; 6 dB). However, previous reports did not mention any differences in MDI-performance for normal hearing and hearing-impaired subjects (section 4.5.3 and Chapter 5), whereas the results for the two groups clearly differed in this manuscript. The set up of this experiment differed in some respects from previous reports. Experiments were carried out at higher sensation levels (1), spectral properties of the carriers were different (2) and experiments were conducted using different set-ups (3).

1. It is unlikely that the sensation level contributed to the differences between this study and others. For normal hearing subjects, the sensitivity to SAM increases when the sensation level increases (LME, p-value < 0.012 for all ISIs) but the difference, given by MDI, was generally similar across sensation levels.

2. MDI is known to decrease with spectral distance (Mendoza et al., 1995b). However, the smaller spectral differences between target and masker in this study are unlikely to have affected the difference between the results from Chapters 5 and 8. They may affect the sensitivity to SAM differently for normal hearing and hearing-impaired subjects due to the difference in spectral resolution. However, this will affect the sensitivity to SAM to a similar extent when presented in unmodulated masker and modulated maskers. It is unlikely that these effects would affect the difference for hearing-impaired subjects, especially since Lorenzi et al. (1997) measured modulation masking using similar carriers, and reported a similar reduction in sensitivity to SAM for normal hearing and hearing-impaired subjects.

3. The results described in this manuscript were conducted using a 1I-2AFC set-up to avoid:

    (a) one of the intervals becoming an additional reference, depending on the duration between the intervals.

    (b) adaptation effects from the preceding pair of stimuli. Adaptation effects have been found lasting for ISIs as long as 625 ms, which suggests that, in order to prevent adaptation from occurring, at least several seconds should be used to separate the intervals. This changes the task of a subject from detecting a difference between two intervals (2/3 AFC) to an absolute judgment of modulation.

These points are unlikely to have affected the differences between the groups since they would mainly influence the overall sensitivity and not MDI.

As has previously been reported by Bacon and Moore (1993), non-modulated maskers reduced the sensitivity to SAM relative to the condition without maskers. This is attributed to masker energy falling in the auditory filters excited by the target, leading to a reduced performance. An alternative way of representing interference is the sensitivity to SAM presented in maskers with respect to the sensitivity to SAM without maskers (see Figure 9.7).

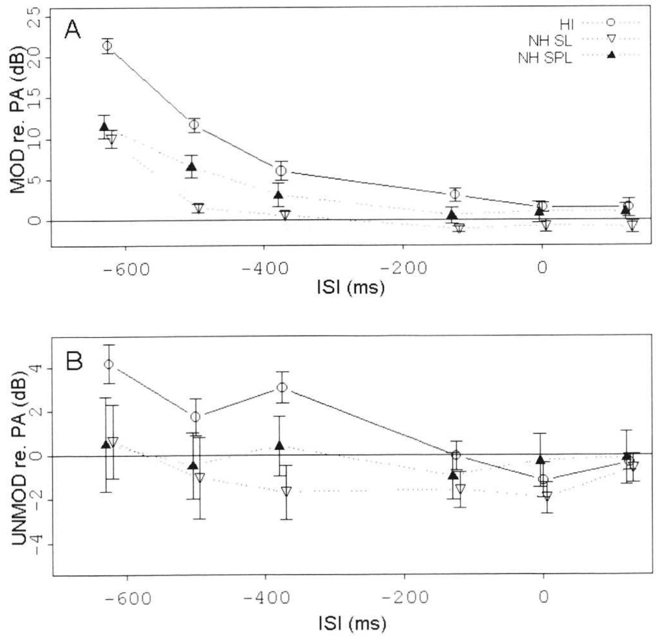


Figure 9.7: *Similar to Figure 9.4, but now with the sensitivity in the presence of a modulated masker (9.7A) and non-modulated masker (9.7B) given relative to the sensitivity without a masker.*

Figure 9.7A shows the reduction in sensitivity to SAM due to modulated maskers for hearing-impaired subjects (circles) and normal hearing subjects (triangles). Three asymptotic functions of ISI are given. For both normal hearing and hearing-impaired subjects, interference decreases rapidly as ISI increases. However, care should be taken when interpreting the data from some hearing-impaired subjects. These subjects were unable to complete the task for simultaneous presentation of teh target and the fully modulated masker. A more shallow decreasing interference is found for normal hearing subjects measured at 50 dB SL with respect to normal hearing subjects measured at 25 dB SL. Hearing-impaired subjects showed a significant interference for all ISIs (t-test, p-value < 0.01) except for ISI=0 ms (p-value $= 0.1$). For normal hearing subjects, a significant interference was found at 25 dB SL for an ISI of -625 ms (t-test, p-value < 0.01), while the amount of interference at 50 dB SL is significant at -625 ms and at -500 ms (t-test, p-value < 0.01). Figure 9.7B shows the reduction in sensitivity to SAM caused by adding non-modulated maskers. Hearing-impaired subjects show a significantly reduced sensitivity to SAM for simultaneously presented signals (ISI = -625 ms; p-value < 0.01) but not for other ISIs, compared to the situation without maskers. For normal hearing subjects measured at 50 dB SL, the sensitivity to SAM was not significantly altered due to non-modulated maskers. At 25 dB SL, normal hearing subjects showed a significantly increased sensitivity to SAM at ISIs -375 ms and -125 ms due to adding non-modulated maskers. Apparently, the non-modulated 'masker' may serve as an additional reference to improve the modulation detection of the target. This may lead to an underestimation of the differences between the two groups. Figure 9.7 also shows that, although the sensitivity restores quickly as the ISIs increase, the sensitivity to SAM is significantly reduced by 1.4 dB (p-value $= 0.005$) at an ISI of 125 ms for modulated maskers for hearing-impaired listeners. This indicates that the interference for non-simultaneously presented signals, the adaptation effect does not only hold for similar carrier frequencies, but also for carriers that are widely separated.

9.4.2 Factors influencing perceptual grouping

It has been suggested that MDI may be based on the difficulty of hearing out the target modulation as a result of grouping of the modulated components of target and masker. In this case altering factors that are known to degrade auditory grouping will decrease the amount of interference. Figure 9.5 shows four factors that are known to influence perceptual grouping. The first condition (9.5A) was measured using an ISI of -375 ms, partly simultaneous, partly non-simultaneous. The other three figures (9.5B, C, D) are based on non-simultaneous presentation of target and masker, using an ISI of 125 ms. If it is postulated that simultaneous and non-simultaneous MDI reflect the same process, as suggested by the gradual decreased sensitivity to SAM as a function of the ISI, the influence of perceptual grouping on both can be investigated. In the rest of this chapter, both tasks will be referred to as interference. The causes of this interference may be twofold:

1. adaptation, which is a physiological process describing a reduction in neural firing due to pre-exposure to SAM.

2. a more cognitive process called perceptual grouping, reflecting the inability to hear out the target modulation due to the grouping of different components.

The data do not support the idea that perceptual grouping is a possible cause of MDI (see Figure 9.5). Hearing-impaired subjects, do not benefit from a difference in phase between the target and masker modulation (Figure 9.5A). In favor of perceptual grouping, there is a tendency for less interference for different carrier frequencies or modulation rates (Figure 9.5B and 9.5D, respectively).

The last parameter, the modulation depth of the adaptor (Figure 9.5C), was introduced to study the contribution of the modulation strength of the adaptor MDI. A clear effect of non-simultaneous MDI was found for a masker with a modulation depth of 1, whereas a modulation depth of 0.5 for the masker did not result in a significantly reduced sensitivity to SAM compared to the condition preceded by a non-modulated masker. Apparently, non-simultaneous interference only occurs when the modulation depth of the masker exceeds 0.5, which makes it difficult to explain non-simultaneous interference in terms of adaptation. However, the effects are rather small and further research is needed to shed additional light on adaptation as a possible contributor to the observed effect. In favor of adaptation leading to

MDI, the sensitivity to SAM is not affected by a difference in phase for the envelope of the target and masker. The effects of the difference in modulation rate and carrier frequency can be attributed to adaptation by presuming a modulation filter bank for different auditory filters, as suggested by Dau et al. (1997b).

9.4.3 The relationship between release of masking for speech and sensitivity to SAM

One hypothesis for the relation between modulation perception and the release of masking for speech is that the reduced benefit from fluctuations in a background noise for hearing-impaired subjects, is caused by an excessive amount of modulation masking. Since a small number of subjects participated in these experiments, rank correlation coefficients have been determined.

This hypothesis can be tested using the correlation between MDI, as a representation of the amount of masking by modulated maskers, and the masking release for speech. Table 9.4 shows rank correlation coefficients between the masking release for speech (MRS) and MDI for different ISIs. The rank correlation coefficients for simultaneous presentation of masker and target (ISI = -625 ms) are significant (p-value < 0.01). Figure 9.8 indicates that a reduction in masking release for speech can, to some extent, be linked to an excessive amount of masking by the modulated maskers. Other ISIs did not result in significant correlation coefficients[‡]. However, there is a lot of scatter and there appears to be a clear subdivision in two groups based on Figure 9.8.

However, MDI may not be the best way to present the masking produced by modulated maskers since the non-modulated masker (carrier) may cause additional masking or serve as a reference, as is shown in Figure 9.7B. An alternative way of expressing the effect of masking modulations, is given by the difference in the sensitivity to SAM without maskers with regard to the sensitivity with modulated maskers (Figure 9.7A). Correlations between this alternative definition of MDI (MDI_{alt}) and the masking release for speech are also given in Table 9.4.

Figure 9.9 shows scatter plots for some of the conditions given by the last

[‡]Chapter 8 indicated relationships between adaptation and the masking release for speech, which were statistically significant. However, recall that in Chapter 8 target and masker were identical in carrier frequency, leading to larger effects.

	Masking Release for Speech			
	re. eq. SL	re. eq. SPL	re. MDI_{alt} eq. SL	re. MDI_{alt} eq. SPL
MDI_{-625}	-0.69^{**}	-0.69^{**}	-0.74^{**}	-0.79^{**}
MDI_{-500}	-0.46	-0.04	-0.50^{*}	-0.46
MDI_{-375}	-0.28	-0.42	-0.62^{*}	-0.55^{*}
MDI_{-125}	-0.42	-0.43	-0.51^{*}	-0.31
MDI_{0}	-0.21	-0.43	-0.39	-0.14
MDI_{125}	-0.01	-0.06	-0.41	-0.17

Table 9.4: *Rank correlation coefficients were determined based on the masking release for speech and MDI for an ISI given as subscript. SL and SPL stand for the exposure levels at which the experiments are conducted (resp. 25 and 50 dB SL) Asterisks denote the level of significance (*p-value < 0.05 *; p-value < 0.01 **).*

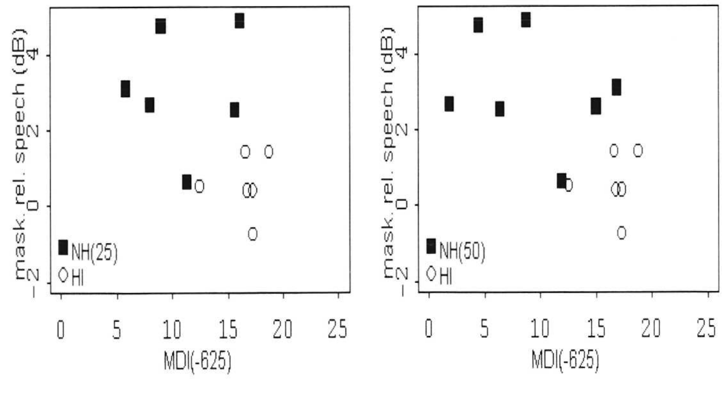


Figure 9.8: *Scatter plots of the masking release for speech and MDI for simultaneous masking. Exposure level for normal hearing subjects is 25 dB SL (left) or 50 dB SL (right). Hearing-impaired subjects are presented by circles, normal hearing subjects by squares.*

two columns of Table 9.4 that correlated with the release of masking for speech. The normal hearing data (squares) are either obtained by measuring at the same sensation level (25 dB SL) as the hearing-impaired subjects (circles) which are given on the left, or at 50 dB SL (right panel). Subplots on the conditions achieved at 50 dB SL at the second and fourth row did not result in significant correlations

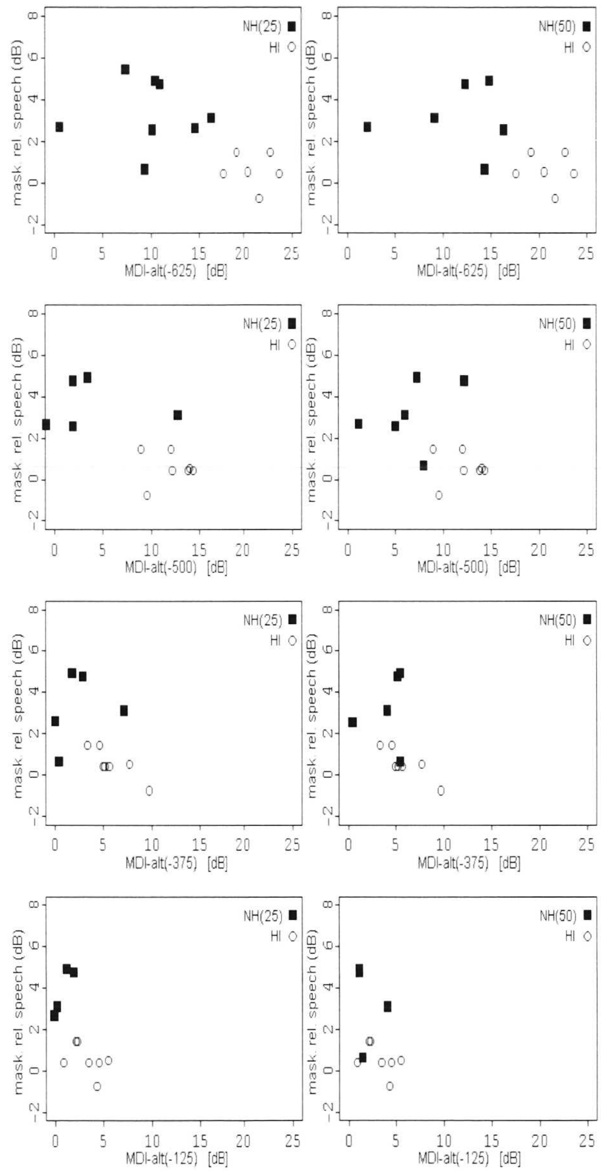


Figure 9.9: *Similar to Figure 9.8, but instead of MDI, MDI_{alt} (the reduced sensitivity to SAM due to modulated maskers relative to the condition without maskers) is used. Data for ISIs -625, -500, -375 and -125 ms are given by the first, second, third and fourth row, respectively.*

with the masking release for speech. Other subplots all indicate that a reduced masking release for speech coincides with an excessive masking of SAM. However, there is is a lot scatter and little overlap between the data from normal hearing and hearing-impaired subjects, which makes the predictive power relatively poor.

9.5 Conclusions

This study can be divided into three parts. The first part studied the sensitivity to SAM as a function of the ISI between masker and target. The results indicate that:

1. in contrast to other studies, large differences are found in the sensitivity to SAM due to modulated maskers between normal hearing and hearing-impaired subjects.

2. there are similarities for interference for simultaneous (MDI) and non-simultaneous presentation adaptation). As function of increasing ISI, a gradual reduction in sensitivity to SAM was found. This may either reflect a decreasing effect of grouping cues, or indicate that classical MDI (Yost and Sheft, 1989) can at least partly be attributed to adaptation.

3. interference for non-simultaneously presented signals is reduced, but still observed for carriers that are separated by 1 octave.

The second part focused on factors indicating whether interference occurred based on perceptual grouping or on adaptation. These effects cannot be segregated based on:

1. a similar interference for experiments with envelopes of masker and target 180^0 out of phase or in phase.

2. identical carrier frequencies leading to similar interference for carrier frequencies that are separated by one octave.

3. identical modulation rates leading to more interference.

4. the fact that a larger modulation depth of the adaptor leads to more interference.

The third part focused on the reduced masking release for speech and the relationship with an excessive amount of modulation masking. Significant rank correlation coefficients are found indicating that masking release for speech can be attributed to an excessive amount of masking of SAM. There is however a lot of scatter in the data. This may still imply that a reduced masking release for speech can be attributed to adaptation, but that this adaptation is carrier specific. An experimental set-up that is more in line with modulation masking (Houtgast, 1989), by using wideband carriers for the target and the masker, may provide additional information.

Chapter 10

Final discussion

10.1 Measuring SAM perception

Four different experiments that focus on the perception of SAM are described in this thesis. Modulation perception was assessed using three different paradigms. Subjects were asked: to detect a difference in the modulation depth (modulation discrimination; Chapter 5); to match the modulation depth of a target to a given reference depth (modulation matching; Chapter 6), or to detect the modulation of a carrier (modulation detection; Chapters 8 and 9). The experiments are visualized in Figure 10.1. The parameters involved and the Chapters in which they are examined are summarized in Table 10.1.

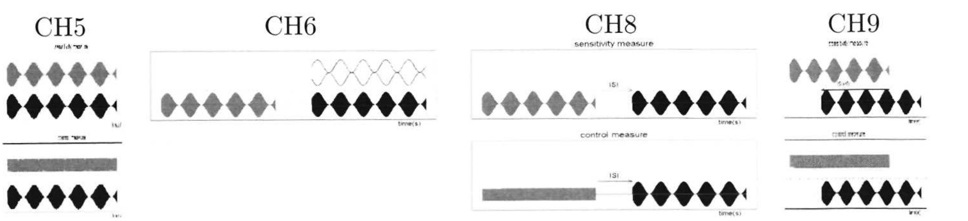


Figure 10.1: *Four conditions in which the sensitivity to SAM was assessed. The task was to detect the modulation of a target (black) in the presence of a masker (gray), that was modulated (upper panel) or non-modulated (lower panel) or to equalize the modulation depth.*

	CH 5	CH 6	CH 8	CH 9
Paradigm	discrimination	matching and detection	detection	detection
Procedure	3AFC	n.a.	1AFC	1AFC
Hearing	NH and HI	NH and HI	NH and HI	NH and HI
m_{target}	m_{ref} = 0:0.18:0.30	m_{ref} = 0.5:0.7	m_{ref} = 0	m_{ref} = 0
m_{masker}	m = 0;0.18;0.30	n.a.	m = 0;1	m = 0;0.5;1
Modulation rate	4:8:16 Hz	8 Hz	8 Hz	8 Hz
Bandwidth	n.a.	narrowband/wideband	n.a.	n.a.
Carrier (target)	1.4 kHz (tone)	1kHz/4kHz (noise)	1 kHz (tone)	1 kHz (tone)
Carrier (masker)	0.5 kHz; 4 kHz (tones)	1kHz/4kHz (noise)	1 kHz (tone)	2 kHz (tone)
Sensation level	MCL	10/25 dB SL	25; 50 dB SL	25; 50 dB SL
ISI	0 ms	200 ms	0..625 ms	-625..125 ms

Table 10.1: *The parameters considered and the chapters, in which they are examined. The abbreviation n.a. indicates that the situation was not applicable for experiments described in the chapter*

10.1.1 Modulation Discrimination Interference (MDI)

The results from the MDI experiments for hearing-impaired subjects indicate that the amount of interference is often similar to, or smaller than, the interference for normal hearing subjects. Any differences in MDI can be attributed to a higher sensitivity to SAM for hearing-impaired subjects when the target is presented in the presence of non-modulated maskers, rather than the sensitivity to SAM in the presence of modulated maskers.

10.1.2 Modulation matching

In modulation matching experiments, the modulation depth of a target has to be adjusted until it sounds just as modulated as a given reference depth. In these experiments, target and reference differed in bandwidth, center frequency, or sensation level and experiments were conducted for hearing-impaired subjects and normal hearing subjects (Chapter 6). Changing the bandwidth alters the perception of the modulation depth. This is mainly due to the difference in inherent fluctuations determined by the bandwidth of the carrier. Hearing-impaired subjects tended to adjust the modulation depth of the target to be higher than the reference depth.

10.1.3 Adaptation

The finding that hearing-impaired subjects generally adjust the modulation depth of a target to be higher than the modulation depth of the reference (Chapter 6) suggests that pre-exposure to SAM may reduce the sensitivity to SAM. These findings have been confirmed in Chapter 8. Measuring modulation detection preceded by a modulated adaptor leads to a reduced sensitivity to SAM with regard to a non-modulated adaptor. This effect, referred to as adaptation, was observed for all hearing-impaired subjects, whereas, in general, normal hearing subjects did not demonstrate this effect.

These measurements indicate that adaptation to SAM occurs for pre-exposure of shorter durations than previously reported (Tansley and Regan, 1979; Tansley and Suffield, 1983; Regan and Tansley, 1979), at least for hearing-impaired subjects. The exact reasons for hearing-impaired subjects being more susceptible to this adaptation to SAM remained unclear within this project. Chapter 8 discussed three possibilities:

1. their auditory system is still analyzing the first signal (adaptor), while being exposed to the second modulated signal (target) and SAM is only detected for modulation depths larger than the 'after image' of the adaptor.

2. physiologically, the AM-filters (Millman et al., 2002) of hearing-impaired subjects may recover more slowly, adapt more quickly, or do both.

3. psychophysically, the temporal focus of the hearing-impaired ear may be inadequate to separate the modulated adaptor and target.

10.1.4 From MDI to adaptation

The sensitivity to SAM in the presence of a modulated or a non-modulated signal was assessed using time discrete intervals ranging from simultaneous (MDI) to clearly non-simultaneous presentation (adaptation). The results indicated that adaptation occurs even when the target and reference differ in carrier frequency. If the target and masker are presented partly simultaneously, the sensitivity to SAM in the presence of a competing modulated signal slowly reduces as the target and masker become more overlapping, until the asynchrony equals 3 periods (375 ms) after which sensitivity abruptly decreases. Similar effects, but to a smaller extent, have been reported for normal hearing subjects. The reduction in sensitivity to SAM at each inter-stimulus interval is larger for hearing-impaired subjects than for normal hearing subjects. even for simultaneous presentation. Hearing-impaired subjects also indicated a slower recovery of the sensitivity to SAM than normal hearings subjects.

10.1.5 Non-linear modulation perception

The results described in Chapters 5 and 9 do not support each other. Whereas Chapter 5 does not report significant differences in MDI between normal hearing and hearing-impaired subjects. Chapter 9 indicates clearly significant differences between the two groups for simultaneous presentation (-625 ms). The conditions in Chapters 5 and 9 differ in a couple of ways. as described in Table 10.1. In Chapter 5. data were obtained using a 3I-3AFC procedure for modulation discrimination (reference depths 0.18 and 0.30). whereas Chapter 9 describes a 1I-2AFC procedure for modulation detection (reference depth 0). It is unlikely that the differences between the chapters can be attributed to the different procedures. The procedure used is likely to have an effect on the overall sensitivity. but not on MDI. as MDI expresses the difference between two conditions examined using the same procedure. Hence. the differences in the results from Chapters 5 and 9 are more likely to have been introduced by the difference in reference depths.

The perception of clearly audible modulation depths was determined using matching experiments (see Chapter 6). The results suggest that the perception of SAM for noise carriers depends primarily on the inherent fluctuations in the signal. The appendix in Chapter 6 showed that the results from the matching experiments of normal hearing subjects can be predicted quite accurately using models described in the literature. These models transform the fluctuations of a signal into a statistical output. The data could be modelled fairly accurately using the statistical output at a given modulation depth. However. the predictive power of the models was clearly improved when the ratio between the statistical output at a given modulation depth and at the modulation detection threshold was used. This is examined in the footnotes to section 6.6.2. Care should be taken when using the results from Chapter 6 to explain the discrepancy between the results from Chapters 5 and 9. Whereas the inherent fluctuations are clearly present for noise carriers. they are absent for pure tone carriers (Chapter 5 and 9). On the other hand, the effect of the difference in inherent fluctuations for the different carriers in Chapter 6 has already been incorporated by determining the statistical output at the modulation detection threshold. Since the predictions of the models use the ratio of statistical outputs, the actual effect of the inherent fluctuations on the predictions may be smaller than originally thought. Hence, the modulation depth alone does not explain the contrasting conclusions in Chapters 5 and 9.

Non-linearities for modulation perception have been described in the literature (von Fleischer, 1980; Ozimek and Sek, 1988; Wakefield and Viemeister, 1990). Weber's law for modulation discrimination holds for modulation depths up to approximately -7 dB (see also section 4.2.5), after which the sensitivity to a change in SAM increases. Indirectly, the reference depths used also determine the modulation range in which perception occurs. This may affect the differences in modulation perception, since relatively sensitive areas may expose differences between groups more easily than less sensitive areas. If modulation perception is correctly described according to $20\log_{10}(m)$, two effects would have been expected in Chapter 6. Since a change in bandwidth induces a difference in the modulation depth perceived, an opposite change in bandwidth should result in the same effect in the opposite direction. In addition, the effects for reference depths 0.5 and 0.7 should be equal in size. The results in Figure 6.6 show that this is not the case. Increasing the bandwidth indeed gives similar results for the two reference depths (parallel lines). However, if the bandwidths are decreased, the two lines converge. This suggests that modulation perception does not occur according to $20\log_{10}(m)$.

The conflicting results from Chapters 5 and 9 may be explained in terms of the nonlinear behavior for modulation perception. Assume that the statistical output is related to SAM perception, as in Figure 6.7. Since the curve relating the statistical output to the modulation depth for highly modulated signals is given by a steep slope at these high modulation depths, the conflicting results from Chapters 5 and 9 may be explained. This steepness causes the differences in modulation depth perception, expressed in dB, to become relatively small. Using $20\log_{10}(m)$ to determine the modulation depth would therefore result in smaller differences for high modulation depths than for lower modulation depths. Hence, a lack of differences between both groups in Chapter 5 may be due to the fact that the differences in modulation depth were detected at areas in which SAM perception is relatively insensitive. At lower modulation depths (as in Chapter 9), modulation perception is described by a relatively sensitive area (less steep), causing the differences for both groups to be emphasized in Chapter 9. Hence, the sensitivity to SAM may be reduced for hearing-impaired subjects with regard to normal hearing subjects for simultaneous and non-simultaneous presentation of target and masker.

10.2 Speech intelligibility and fluctuating noise

A number of studies (Noordhoek et al., 2001; Festen and Plomp, 1983; Dreschler, 1983) have offered a good insight into the psychoacoustical tasks that contribute to intelligibility of speech. Most studies focussed on impaired intelligibility in a continuous background noise. The results indicated that measures of spectral resolution are strongly related to speech intelligibility, whereas measures of temporal resolution generally result in weak correlation coefficients. As Noordhoek et al. (2001) pointed out, audibility is an essential parameter limiting the explained variance within such correlation studies. To avoid these problems, speech can be presented at relatively high levels, or speech intelligibility can be expressed in the SII, or determined for subjects with relatively flat hearing losses.

The study described in this thesis, focussed on the improvement in speech perception in fluctuating rather than continuous noise. In this study, the intelligibility of speech was determined using a continuous and a fluctuating background noise keeping the audibility of speech roughly constant. The difference in SRT for both measurements is termed the masking release for speech. The intelligibility in a fluctuating noise may increase due to temporary improvements in S/N ratio. This is known as 'glimpsing'. These improvements in S/N ratio may not assist subjects who are unable to use these improvements due to a limited resolving power of their ears.

Masking release for speech may be reduced by limitations in the resolving power of the ear. Chapter 4 divided measures of the resolving power of the ear into three parameters:

1. compressive non-linearity: it is difficult to determine the lack of an effect of compression on intelligibility. Fu and Shannon (1998) determined the intelligibility of bandpass filtered CVC-words (4 bands) as a function of the

expansion applied. Optimal intelligibility was found for the linear condition and performance decreased as expansion increased. Hence, increasing the differences between the peaks and valleys within the signal did not improve the intelligibility of speech. However, enlarging the peaks of the fluctuating noise using a lack of compressive non-linearity in the cochlea may have increased the non-simultaneous masking of speech and therefore decreased the masking release for speech.

2. frequency selectivity: reduced spectral contrast may result in less masking release for speech. Since noise bands may modulate independently, the opportunity to use a high S/N ratio may be determined by the spectral resolution. From a spectral point of view, increasing the number of bands will increase intelligibility, until the bands are as wide as the critical bandwidth. Note that the fluctuating noise barely affected the inter-band coherence, because the signal was only split in two bands. Hence, the broader filters for hearing-impaired subjects may not have influenced the results of the SRT, leaving the masking release for speech unaltered.

3. temporal resolution:

   (a) non-simultaneous masking: as suggested by Festen (1993), the intelligibility of speech in a fluctuating noise is limited by forward masking causing the masking release for speech to decrease.

   (b) gap-detection: the effect of gap-detection on masking release for speech is unclear. Studies relating gap-detection to speech intelligibility in noise (Dreschler and Leeuw, 1990) showed relatively modest correlation coefficients, suggesting that masking release for speech may be unaffected by reduced gap detection

   (c) SAM perception: this thesis assessed the underlying relationship between SAM perception and the masking release for speech. SAM perception without maskers was not found to be related to speech intelligibility. However, masking release for speech strongly correlated with the reduced sensitivity to SAM due to modulated components, which were presented non-simultaneously (Chapter 8) or (partly) simultaneously (Chapter 9).

Apart from a larger MDI for hearing-impaired subjects for simultaneously presented signals, hearing-impaired subjects also experienced more adaptation when the target and masker were presented partly non-simultaneously. Based on the importance of the envelope for speech, this reduced sensitivity may partly explain the lack of masking release for hearing-impaired subjects.

10.2.1 Method of correlational studies

The experiments described in Chapters 7, 8, and 9, used correlation coefficients to identify the underlying relationships between speech intelligibility and measures of modulation perception. These correlations provide, on their own, no justification for a direct link between speech intelligibility and modulation perception. These methods rely on the fact that a reduced performance on one task (modulation detection or discrimination) is related to a reduced or increased performance in another task (masking release for speech). When studying these kinds of relationships between variables, pooling normal hearing and hearing-impaired subjects is useful since it increases the number and range of observations. The validity rests on the assumption that both groups employed the same strategy and the observations were drawn from the same population. However, if the combined groups represent a bimodal distribution rather than a continuum, the correlation coefficient may be overestimated. To verify whether this is the case, it is essential to present scatter plots (*e.g.* Figure 8.5). In this thesis, the two groups were pooled when the data from both groups displayed a continuum and rank-correlation coefficients also indicated significant relationships. Rank correlation coefficients are less sensitive to bimodal distributions, since they do not take the size of the effects into account and have, therefore, been used to check the level of significance reported throughout this thesis.

10.3 Clinical applications

Clinical applications of new psychoacoustical effects may focus on two different areas:

1. diagnosis: Noordhoek et al. (2001) indicated that, in general, the difficulties experienced for intelligibility in noise can be attributed to a reduced spectral or temporal resolution. Although the adaptation effects described in this thesis, may result from reduced temporal resolution. Temporal resolution is often assessed for much shorter observation intervals. This would indicate that a third group could be defined, that have problems with the perception of amplitude modulated signals. If this were true, one could discuss the appropriate measurements to diagnose this group. If the reduced sensitivity to SAM is related to intelligibility, as suggested by our results, one might consider measuring the intelligibility of CVC-words presented in a fluctuating noise with the pre-exposure time to the modulated noise as the dependent variable.

2. hearing-aid algorithms: the distorted modulation perception described in Chapters 8 and 9, is likely to affect speech intelligibility. Modern hearing aids use fast automatic non-linear algorithms, which are known to affect the modulation behavior of incoming signals. Figure 9.5 suggests that adaptation to SAM occurs for pre-exposure to a fully modulated signal, whereas it disappeared for a modulation depth of 0.5. If this effect is confirmed for larger groups, this may have its consequences for the development of new hearing aid algorithms.

10.4 Future research

This thesis has shown that adaptation to SAM occurs for much shorter pre-exposure times than originally thought (Tansley and Regan, 1979). It is striking that this effect is found in all hearing-impaired subjects, but only one normal hearing subject. The process underlying this adaptation effect has remained unclear and further research is needed. Two potential factors have been investigated in more detail in Chapter 9:

1. Perceptual grouping: depending on similarities within the signals (modulation pattern), signals are more or less likely to be fused into a single percept.

2. Adaptation: pre-exposure to SAM reduced the sensitivity to SAM as a result of reduced neural activity after prolonged stimulation (Coombs and Fay, 1985).

In order to examine, which factors are most likely to cause a reduced sensitivity to SAM, several parameters were considered and the results are described briefly in section 9.3.3 and may be useful in further research:

1. tuning for modulation rates: modulating the envelope using narrowband noises would lead to a reduced sensitivity to SAM in the case of adaptation, whereas perceptual grouping would barely occur.

2. dependence on carrier-frequencies; using a multiple tone complex with a harmonic relation as an adaptor without the target frequency is likely to affect perceptual grouping, whereas adaptation may be unaffected when the fundamental frequency is sufficiently large to prevent within channel processing from occurring. However, as indicated by the effect of a modulated carrier at 2 kHz on the sensitivity to SAM at 1 kHz, modulation perception must occur by fusing different filter-outputs simultaneously before the statistical output device. This is illustrated in Chapter 6, where modulation perception was mainly determined by the common bandwidth, and not by the number of filters excited. Hence, a reduced sensitivity to SAM, may also simply reflect an across-channel effect for modulation perception.

3. phase-relations between envelopes; adaptation is likely to be unaffected by differences in the envelope phase of the target and masker, since it is linked to the activity of one channel. Perceptual grouping is likely to be reduced, since modulation phase differences are known to affect perceptual grouping.

4. the effect of modulation depth of the preceding stimulus; adaptation is likely to be reduced as the modulation depth of the adaptor decreases, whereas perceptual grouping may be less affected, since grouping parameters remain roughly constant.

5. within and across channel processing; if the sensitivity to SAM of a pure tone carrier is measured after pre-exposure to a modulated broadband adaptor, the reduced sensitivity to SAM probably cannot be linked to perceptual grouping.

This is similar to using a broadband signal as a target and a pure tone carrier as adaptor.

Clearly a lot of work needs to be done in order to understand the underlying auditory processes that lead to the reduced sensitivity to SAM for hearing-impaired subjects (Chapter 8 and 9). The high correlation coefficients between the reduced sensitivity to SAM due to pre-exposure to a modulated adaptor and masking release for speech suggest that focussing on this topic may not only increase the knowledge regarding the processes underlying MDI, but may also increase the knowledge on speech intelligibility in fluctuating noise.

Chapter 11

Summary

The work described in this thesis can be divided into three topics:

1. Signal properties of speech (Chapters 2 and 3)
2. The perception of sinusoidally amplitude modulated (SAM) signals (Chapters 4, 5, and 6)
3. Reduced sensitivity to SAM due to modulated maskers in relation to the masking release for speech (benefit due to fluctuations in the noise relative to a continuous background noise) (Chapters 7, 8, and 9)

1. Signal characteristics

Chapter 3 describes the development of a database of sounds, based on their spectral and temporal behavior. Although the structure of the speech signal is fairly constant, in terms of both spectral (LTASS) and temporal behavior (modulation spectrum) regardless of language or speaker (see Chapter 2), the modulation characteristics of speech serves as a better identifier of whether noise or speech is presented (Chapter 3). In addition, the effect of the modulation strength of the masker was shown to be the major factor contributing to the differences in speech intelligibility for normal hearing subjects and hearing-impaired subjects. Whereas normal hearing subjects clearly benefitted from the temporal gaps, hearing-impaired subjects could not.

2. SAM perception

The left panel of Figure 11.1 describes the condition by which modulation discrimination interference (MDI; see Chapter 5) can be determined. The target and masker differ in carrier frequency. The task was to detect a difference in the modulation depth of the target. In order to determine the effect of the modulated maskers, MDI was defined as the difference in sensitivity to a change in SAM in the presence of modulated (upper panel) and non-modulated maskers (lower panel). The sensitivity to a change in SAM was severely reduced when modulated maskers were added, but was not significantly different for normal hearing and hearing-impaired subjects. Since differences in modulation depth were detected for relatively highly modulated stimuli (using reference depths of 0.18 and 0.30), the lack of differences between normal hearing and hearing-impaired subjects may be partly attributed to a reduced detection space for such highly modulated stimuli. Hence, the results raised questions on how SAM is perceived for highly modulated signals.

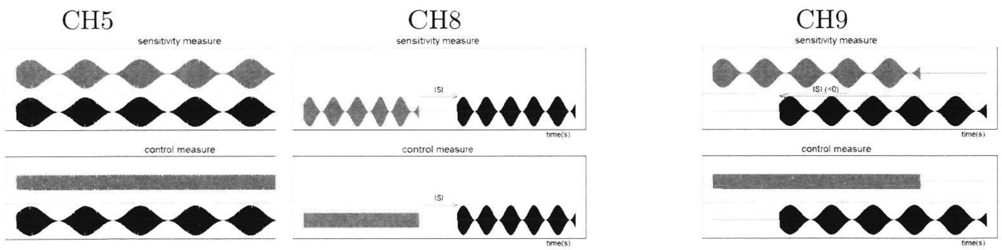


Figure 11.1: *Three conditions used to assess the sensitivity to SAM. The task was to detect the modulation of a target (black) in the presence of a masker (gray), that was modulated (upper panel) or non-modulated (lower panel).*

Experiments were carried out using noise carriers with modulation depths presented well above the threshold of modulation detection (Chapter 6). Subjects were asked to adjust the modulation depth of a stimulus until they perceived it as equally modulated as a reference stimulus. This reference signal differed either in bandwidth, center frequency, or sensation level from the target. For normal hearing subjects, the sensation level had little effect on modulation perception, but bandwidth and center frequency, with the absolute bandwidth of the carrier as the common factor, indicated that the altered modulation perception can mainly be attributed to the bandwidth of the stimuli. The modulation depth of a stimulus with a great amount of slow inherent fluctuations (narrowband or low center frequency) was adjusted higher in modulation depth than a signal with a relatively small amount of slow inherent fluctuations (broadband or high center frequency). The results from hearing-impaired subjects can, for a constant carrier frequency, also be described in terms of the differences in the amount of slow inherent fluctuations of the signals for target and reference. However, there was an unexplained overall trend for hearing-impaired subjects to adjust the modulation depth of the target higher than modulation depth of the reference. Differences in the growth of loudness did not correlate strongly with the differences in the modulation depth for target and reference. This indicates that the differences in physical modulation depths for an equal modulation perception cannot be explained in terms of lack of compression (the magnifying factor for modulation between the conditions).

3. Sensitivity to SAM related to MRS

Normal hearing subjects often experience fewer difficulties when listening to talkers in a fluctuating background noise (music, multiple talkers) than when talkers speak at the same S/N ratio in a continuous background noise (motor). Normal hearing subjects can use the sudden increments in S/N ratio and the context to optimize performance. Figure 11.2 presents a visual example. Although the letters cannot be distinguished from the black bars (right panel), creating white spaces, which are sufficiently large, enables the text to be read. The left panel is an example of text in a continuous background noise. Although the background is rather dark, the text is still readable.

For hearing-impaired subjects intelligibility is comparable when speech is presented at the same signal-to-noise ratio in a continuous or in a fluctuating noise. In Chapter 4, it was discussed that most hearing-impaired subjects experience reduced spectral resolution. To some extent, this may be compared to a reduced contrast*. A transparent piece of paper is included in this thesis. If this piece of paper is placed loosely over Figure 11.2, the left panel becomes difficult to read, whereas the right panel can still be read without any problems. Hence, a reduced

*Although using different colors is a more logical approach, the reduced contrast illustrates the reduced spectral resolution by blurring out the image. Differences in black and white become less apparent, similar to the spectral differences for hearing-impaired subjects.

spectral resolution may be even more detrimental to speech presented in a continuous than in a fluctuating background noise.

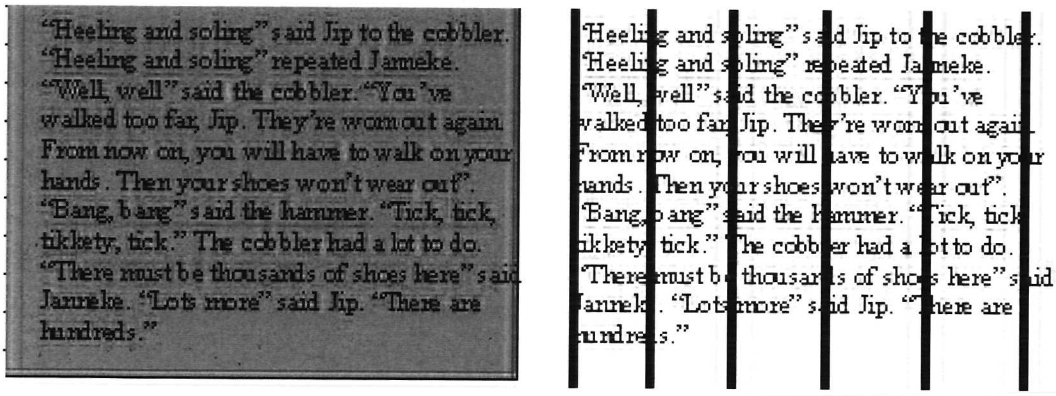


Figure 11.2: *Visual analogue for speech intelligibility presented in a continuous (left panel) and fluctuating (right panel) noise.*
The text was adapted from Jip en Janneke (Schmidt, 1954), with permision from the publisher

The tendency of hearing impaired subjects to adjust the modulation depth of the target to be larger than the modulation depth of a given reference (Chapter 6) may affect speech intelligibility in a fluctuating noise. This was the main motivation for the study described in Chapter 8. Modulation detection thresholds were measured for a pure tone carrier preceded by a pure-tone adaptor that was either non-modulated or fully modulated (see middle panel Figure 11.1). Recovery times (ISIs), ranging from 10 to 625 ms, were used to determine differences in the sensitivity to SAM due to a modulated signal that preceded the target for normal hearing and hearing-impaired subjects. Clear differences in sensitivity to SAM were found for hearing-impaired listeners when a modulated, instead of a non-modulated signal preceded the target. This adaptation to a modulated adaptor indicates that hearing-impaired subjects experience modulation detection interference for non-simultaneous signals.

Chapter 8 reported large differences between the sensitivity to SAM when a target followed an adaptor for normal hearing and hearing-impaired subjects, while Chapter 5 reported barely any differences for simultaneously presented signals using spectral differences to separate target and masker. Therefore, a final study was conducted (Chapter 9, see right panel Figure 11.1), which used different carrier frequencies for target and masker. The target and masker were presented at discrete time intervals ranging from -625 ms (simultaneous) to 125 ms (non-simultaneous). Large differences in the reduced sensitivity to SAM were found for normal hearing and hearing-impaired subjects for simultaneous presentation of target and masker. For both, normal hearing and hearing-impaired subjects, the reduction in sensitivity to SAM was an inverse function of ISI, with most masking for simultaneous presentation tending to zero for larger ISIs. This indicated that normal hearing and hearing-impaired subjects showed similar results as in Chapter 5, due to the large differences in modulation sensation for high modulation depths (modulation discrimination for reference depths 0.18 and 0.30), whereas the differences in modulation sensation are relatively small for lower modulation depths (modulation detection).

Based on the importance of the temporal characteristics to speech (Chapter 3 and 7), the reduced sensitivity to SAM may have its impact on intelligibility. In Chapter 8 and 9, parameters reflecting SAM perception were compared to the improvement in intelligibility when speech was presented in a fluctuating rather than a continuous background noise. The larger the adaptation effect, the smaller the benefit of fluctuations in the noise. It appears that masking release for speech, most likely resulting from 'glimpsing', is reduced by the modulations in the fluctuating masker preceding the target. This can also be illustrated by Figure 11.2. If the part of the piece of transparent paper with dark bars is held over the text, the text is much harder to read. The dark bars represent the reduced sensitivity to SAM, with the shading symbolizing the reduced sensitivity.

Verdomd interessant,

maar gaat U verder ...

Wim T. Schippers

Chapter 12

Samenvatting

Dit proefschrift richt zich op drie vraagstukken;

1. Hoe kan spraak gekarakteriseerd worden ? (hoofdstuk 2 en 3).

2. Hoe worden SAM-signalen* waargenomen bij normaal horenden en slechthorenden ? (hoofdstuk 4, 5 en 6)

3. In hoeverre kan de verminderde MRS† bij slechthorenden worden toegeschreven aan de verminderde gevoeligheid voor SAM-signalen in aanwezigheid van andere SAM-signalen ? (hoofdstuk 7, 8 en 9)

1. Eigenschappen van spraak

Onafhankelijk van spreker en taal, zijn signaal eigenschappen zoals het spectrum (LTASS) en het gedrag over tijd (modulatiespectrum) relatief constant (zie hoofdstuk 2). Hoofdstuk 3 beschrijft hoe een grote database van ruizen is gecreëerd en geanalyseerd op hun temporeel en spectraal gedrag. Spraak bleek het meest uitgesproken te zijn in haar temporeel gedrag dat daarom een goede mogelijkheid biedt om spraak te scheiden van andere signalen. Verder is de sterkte van de omhullende van het signaal de belangrijkste factor die de verschillen in

*Signalen waarbij de intensiteit van het signaal verloopt zoals een sinus (als functie van de tijd)
†*Masking release for Speech*; Ofwel het verbeterd spraak verstaan in een fluctuerende ruis t.o.v. een continue ruis

spraakverstaan voor normaal horenden en slechthorenden bepaald. Terwijl het spraakverstaan voor normaal horenden beter wordt wanneer er gaten in de ruis zitten, kunnen slechthorenden hier geen gebruik van maken.

2. SAM perceptie

Het linker deel van figuur 12.1 beschrijft een conditie om modulation discrimination interference (MDI) te meten (hoofdstuk 5). De maskeerder en het testsignaal verschillen in grondfrequentie van elkaar. Het doel van het experiment is om een verschil in modulatiediepte waar te nemen (de sterkte van de modulatie). Om het effect van het moduleren van de maskeerders te bepalen, wordt de gevoeligheid voor SAM bepaald in aanwezigheid van gemoduleerde maskeerders (linksboven figuur 1) en ongemoduleerde maskeerders bepaald (linksonder figuur 1). De gevoeligheid voor een toename in modulatiediepte nam drastisch af door het toevoegen van gemoduleerde maskeerders, maar was niet significant verschillend voor slechthorenden en normaal horenden. Omdat discriminatie van modulatie diepte over het algemeen pas gebeurde voor sterk gemoduleerde signalen, zou het gebrek aan verschillen tussen normaal horenden en slechthorenden verklaard kunnen worden in termen van een beperkt meetbereik. Ook kan het zijn dat de perceptie van sterk gemoduleerde signalen verschilt van de perceptie van zwak gemoduleerde signalen.

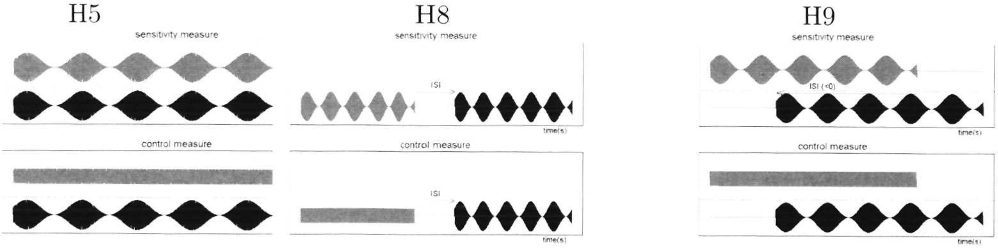


Figure 12.1: *De drie condities gemeten in de verschillende experimenten (hoofdstuk 5, 8, en 9). Het doel was steeds om de modulatie van een testsignaal (zwart) te detecteren in bijzijn van een gemoduleerde maskeerder (grijs). Het effect van de gemoduleerde maskeerder werd bepaald door de maskeerder gemoduleerd (boven) en ongemoduleerd (onder) aan te bieden.*

Om die reden hebben we besloten om experimenten uit te voeren gebruik makend van ruisbandjes waarvan de modulatiedieptes boven de detectiedrempel werden aangeboden (hoofdstuk 6). Hierbij werd aan de deelnemers gevraagd om de sterke van de modulatie zo in te stellen dat ze net zo gemoduleerd klinken als een zeker referentiesignaal. Dit referentiesignaal verschilde van het testsignaal in bandbreedte, centrum frequentie of niveau. De waargenomen modulatie diepte werd niet anders door het niveau te veranderen, terwijl een verandering van frequentie en bandbreedte wel degelijk het modulatie percept veranderden. Dit effect werd voornamelijk bepaald door de bandbreedte van de stimuli. Wanneer de langzame interne fluctuaties toenemen (bij afgenomen bandbreedte) wordt de modulatie diepte van het signaal met veel langzame interne modulaties hoger ingesteld dan het signaal met weinig langzame interne fluctuaties. De resultaten van slechthorenden waren in het algemeen vergelijkbaar met de resultaten van normaal horenden. Slechthorenden hadden echter meestal de neiging om de modulatiediepte van het testsignaal hoger in te stellen dan de modulatie diepte van het referentiesignaal. Verschillen in luidheidopbouw droegen niet bij aan het verklaren van variantie in de data.

3. Relatie tussen temporele resolutie (SAM) en spraakverstaan

Wanneer normaal horenden moeten luisteren in achtergrondruis, is het verstaan vaak veel beter in een fluctuerende omgeving (muziek, andere sprekers, of het kloppen van een hamer) dan in een continue omgeving (geluid van een motor) wanneer deze geluiden even hard worden aangeboden. Normaal horenden kunnen als het ware tussen de gaten door luisteren en kunnen gebruik maken van een kort moment waarin het gemakkelijk is om te luisteren. Bovendien zijn zij in staat de woorden op het moment dat de maskering hoog is zelf in te vullen. Om dit te verduidelijken, wil ik een visuele analogie gebruiken (figuur 12.2; rechts). Hoewel de letters, waar de zwarte balken overheenvallen absoluut onleesbaar zijn, kun je door de witte gaten groot genoeg te kiezen, uit de context opmaken wat er moet staan. Het linker deel van figuur 12.2, is een variant op de continue ruis. Hoewel de achtergrond vrij donker is, is de tekst nog prima te lezen.

Voor slechthorenden maakt het echter niet zo gek veel uit of de spraak nu wordt aangeboden in een fluctuerende ruis of in een continue achtergrondruis. Nu hebben we in hoofdstuk 3 gezien dat de spectrale resolutie (enigszins vergelijkbaar met het contrast‡) minder is voor slechthorenden dan voor normaal horenden. In dit boekje treft u een transparant aan. Wanneer dit losjes over de tekst gelegd wordt, wordt de linker tekst duidelijk minder goed leesbaar, terwijl de rechter tekst gemakkelijk leesbaar blijft. Een verminderde spectrale resolutie is dan waarschijnlijk ook meer bepalend voor de spraak dan voor de ruis en bepaald daarom de overall afname in spraak verstaanbaarheid, welke vergelijkbaar is voor spraak in een continue en fluctuerende achtergrondruis.

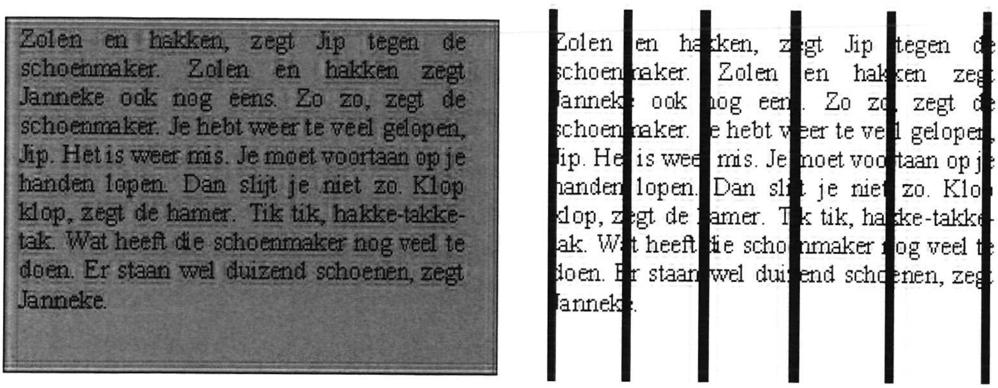


Figure 12.2: *Visuele analogie voor spraakverstaan in een continue (links) en fluctuarende (rechts) achtergrondruis.*
Tekst is afkomstig uit Jip en Janneke (Schmidt, 1954), met toestemming van de uitgever.

De neiging van slechthorenden om de modulatie diepte hoger in te stellen, kan van invloed zijn op het spraakverstaan in een fluctuerende achtergrondruis en verdiende daarom extra aandacht (hoofdstuk 8). De gevoeligheid voor SAM werd gemeten wanneer het testsignaal voorafgegaan werd door een gemoduleerde en ongemoduleerde maskeerder, dit is een continue zuivere toon (zie middelste figuur 12.1) met ISIs van 10 tot 625 ms. Ongeacht of het testsignaal vooraf gegaan werd door een gemoduleerde of ongemoduleerde maskeerder bleef de gevoeligheid constant

‡Hoewel het gebruik van kleuren een meer realistisch beeld zou opleveren, wordt het gereduceerde contrast weergegeven door het vager worden van het beeld. De verschillen in zwart en wit worden minder duidelijk, net zoals de spectral verschillen in geluid voor slechthorenden.

voor de meeste normaal horenden. Slechthorenden bleken echter minder gevoelig voor SAM te zijn wanneer het testsignaal vooraf gegaan werd door een amplitude gemoduleerde maskeerder in plaats van een ongemoduleerde maskeerder.

De vraag die onbeantwoord blijft is, waarom vinden we wel verschillen tussen normaal horenden en slechthorenden bij niet-simultane presentatie (hoofdstuk 8), maar niet bij simultane presentatie (hoofdstuk 5). Om dit uit te zoeken hebben we in hoofdstuk 9 een experiment uitgevoerd waarbij het testsignaal en het maskeer signaal verschillende frequenties hadden, dit is vergelijkbaar met hoofdstuk 5. De tijdsintervallen verliepen van simultaan (ISI=-625 ms, gelijk aan hoofdstuk 5) tot niet-simultaan (zie rechter deel van figuur 12.1 ISI>0; vergelijkbaar met hoofdstuk 8). Zelfs voor simultane aanbiedingen werden nu grote verschillen in de gevoeligheid voor SAM tussen normaal horenden en slechthorenden gevonden. Dit suggereert dat de relatief beperkte meetruimte waarbinnen verschillen in modulatiediepte konden worden waargenomen, ervoor gezorgd heeft dat er geen verschillen gevonden zijn in hoofdstuk 5.

Deze verminderde gevoeligheid voor SAM zou van invloed kunnen zijn op het spraakverstaan in een fluctuerende ruis. Spraak karakteriseert zich door amplitude-modulaties (zie Hoofdstuk 1 en 2). De resultaten van hoofdstuk 8 en 9 bleken dan ook te correleren met het verschil in spraakverstaan in een continue en een fluctuerende ruis; hoe groter de afname in gevoeligheid voor SAM door de gemoduleerde 'maskeerder', hoe kleiner de winst als gevolg van de gaten in de ruis. Om dit weer met de visuele analogie van figuur 12.2 uit te leggen, moet het onderste gestreepte deel van het doordrukvel, naast de balkjes van figuur 12.2 gelegd worden. De balkjes worden steeds donkerder om de afname in gevoeligheid weer te geven, hierdoor neemt de leesbaarheid van de tekst af.

Chapter 13

Mag ik bedanken?

Eindelijk zit het er op! Als ik mijn vriendenkring moet geloven, loop ik al een jaar of zes te vertellen dat mijn promotie volgend jaar wel afgerond zal zijn. Velen geloofden er niet meer in, maar.... Het resultaat van dagenlang programmeren, schrijven, lezen, en analyseren heeft dan uiteindelijk toch een boekje opgeleverd.

Dank ben ik verschuldigd aan Wouter Dreschler, voor de gelegenheid om mijn werk in een promotie te laten eindigen. Maar ook wil ik Wout bedanken door geloof te tonen in de, vaak op het eerste oog, ridicule uitkomsten van mijn experimenten. Bovendien vertrouwde hij in de onfeilbaarheid van mijn resultaten, in het begin vaak meer dan ik zelf. Daarnaast wil ik graag mijn copromotor Tammo Houtgast bedanken, die door het voeren van discussies vaak concepten tot een helderder beeld heeft kunnen brengen. Verder wil ik de leden van de leescommisie prof. dr. E. de Boer, prof. dr. T. Dau*, prof. dr. W.J. Fokkens, prof. dr. A. Kohlrausch, prof. dr. ir. H. Spekreijse, en prof. dr. G.F. Smoorenburg, bedanken voor de tijd, moed en moeite die jullie hebben getoond om mijn proefschrift te bestuderen en verbeteren met intelligente opmerkingen.

De proefpersonen ben ik veel dank verschuldigd. De normaal horenden, veelal tijdelijke passanten van het Audiologisch Centrum, die geheel vrijblijvend (op een enkele traktatie daargelaten) vaak uren in een testcabine naar piepjes hebben zitten luisteren. Nog meer dank ben ik verschuldigd aan de slechthorenden, patiënten uit

*With a special word to prof. T. Dau. Thanks for your time and effort to struggle through this thesis, improving it by your intelligent comments.

de kliniek of van de A.G. Bellschool. Echter bovenal ben ik Rein Stolp dankbaar door zich bereid te tonen om als testcase te dienen voor mijn vaak te lange testen om vervolgens tot een kleinere proefset te kunnen komen.

Simon Wiersma, en later Láslzó Körössy, bedankt voor de (mentale) steun wanneer Delphi weer eens een geheel eigen leven begon te leiden. Natuurlijk, wil ik mijn directe collegae, Niels en Arjo in het bijzonder, bedanken voor de vaak gezellige en prettige momenten. Niek en Sidonne, bedankt voor de fysieke steun in het schrijfproces en de hulp bij het selecteren van proefpersonen. Maar bovenal Bas Franck.... Je houdt het niet jaren met elkaar uit op een, niet altijd even goed geventileerde, kamer zonder een goede band op te bouwen. Mede door je goede grappen ('OUI/WIE') en openheid heb ik mijn tijd op het AMC als uiterst positief ervaren. Daarnaast, heeft jouw eeuwig enthousiasme, met name richting je vrouw en kinderen, een uiterst positieve sfeer geschapen. En nu ik zelf in die situatie zit... Bas, je hebt het me de afgelopen zes jaar niet vaak horen zeggen, maar je hebt helemaal gelijk... Het is echt zo leuk. Bedankt!!!

Jan, Heleen en Maaike, bedankt voor alle steun, liefde en opbeurende woorden. Net als mijn vrienden, wisten jullie vaak de juiste woorden te zeggen om mij met nog meer enthousiasme naar het werk te laten gaan.

Lieve Rebecca, bedankt voor je steun en liefde. Sinds het schrijven van het eerste hoofdstuk is er veel veranderd, maar jij hebt met name veel goeds gebracht. Wat begon als een statistisch vraagje, is tot iets moois uitgebloeid.

Chapter 14

Curriculum Vitae

Jan Koopman werd geboren op 13 november 1974 te Amsterdam. Nadat hij in 1993 het atheneum-diploma behaalde op Damstede (tegenwoordig Bernhard Nieuwentijt College) te Amsterdam kon hij in dat zelfde jaar beginnen met de studie natuurkunde. Deze studie werd in de zomer van 1998 afgerond, met als afstudeerrichting medische/experimentele fysica. Tijdens zijn studie heeft hij stage gelopen bij de afdeling experimentele en klinische audiologie, onder begeleiding van ir. R. van der Horst. Na zijn stage kon hij voor 2 jaar verder gaan binnen de klinische audiologie door aan het project SPACE deel te nemen. SPACE was een samenwerkingsverband van een aantal universtieiten (Nederland : AMC en VU, Amsterdam; Erasmus, Rotterdam; Engeland : Cambridge; Duitsland : Carl von Ossietzky-Universität Oldenburg, Gießen, Georg Simon Ohm Fachhochschule Nürnberg) waarbij gekeken werd naar de problemen van normaal horenden en slechthorenden op de werkplek. Hierna werd hem de mogelijkheid geboden om zijn promotie onderzoek af te ronden, wat heeft geleid tot dit proefschrift.

14.1 Artikelen

Published:

Koopman, J., Franck, B. A. M., and Dreschler, W. A. - Toward a representative set of "real-life" noises. *Audiology* **2001**; 40: 78-91

Abstracts:

Koopman. J.. Plasmans. C.J.P.. and Dreschler. W.A. - Amplitude modulation perception for people with normal hearing. *J. Acoust. Soc. Am.* **2002**: 111: 2470

Koopman. J.. Houtgast. T.. and Dreschler. W.A. - Temporal interference in amplitude modulation perception. *J. Acoust. Soc. Am.* **2002**: 111: 2338

Koopman. J.. van der Horst. R.. and Dreschler. W.A. - Amplitude modulation discrimination interference (MDI) and speech intelligibility for normal-hearing and hearing-impaired subjects. *J. Acoust. Soc. Am.* **1999**: 105: 967

Submitted:

Koopman J., Dreschler. W.A. - Amplitude modulation discrimination interference for normal and hearing-impaired subjects. *Submitted to Ear and Hearing.*

Koopman J.. Plasmans. C.J.P.. Dreschler. W.A. - Modeling the perception of supra-threshold Amplitude modulated signals for normal hearing subjects. *Submitted to J. Acoust. Soc. Am.*

Koopman J., Dreschler. W.A. - The perception of supra-threshold Amplitude modulated signals for normal hearing subjects and hearing-impaired subjects. *Submitted to Ear and Hearing.*

Koopman J., Houtgast, T., Dreschler. W.A. - Adaptation to amplitude modulated signals and its relevance to listening in noise. *Submitted to J. Acoust. Soc. Am.*.

Koopman J.. Houtgast. T.. Dreschler. W.A. - Non-simultaneous masking and its relevance to listening in noise. *Submitted to J. Acoust. Soc. Am.*.

Internal Reports:

Koopman J., Franck, B.A.M., Dreschler. W.A. , Pastoors. A.D., and Kiessling. J. - Strategy for the individual fitting of communication devices in noisy environments. *SPACE final report E3. AMC. SPACE-report RP-E3.* 1999.

Koopman J., Vlaming M.S.M.G.. Goedegebure. A., Verschuure. J., Dreschler, W.A. - Monitoring of the market for digital hearing aids and the implementation of new algorithms in commercial hearing aids. *SPACE report RP-EP1a*, 2000.

Koopman, J. Handleiding bij computerprogramma ten behoeve van de validiteitsbepaling "Horen op het werk. of werken op het gehoor" 1998.

Bibliography

ANSI. 1996. *American National Standards Institute;New York American national standard methods for Specification for audiometers; ANSI S3.6-1996.*

ANSI. 1997. *American National Standards Institute; New York American national standard methods for calculation of the speech intelligibility index; ANSI S3.5-1997.*

Arai, T. and Greenberg, S. 1998. Speech intelligibility in the presence of cross-channel spectral asynchrony. *IEEE/ICASSP, Seattle*, pages 933–936.

Bacon, S. P. 1999. Some effects of background noise on modulation detection interference. *Hear.Res.*, 129:20–26.

Bacon, S. P. and Gleitman, R. M. 1992. Modulation detection in subjects with relatively flat hearing losses. *J.Speech Hear.Res.*, 35:642–653.

Bacon, S. P. and Grantham, D. W. 1989. Modulation masking: effects of modulation frequency, depth, and phase. *J.Acoust.Soc.Am.*, 85:2575–2580.

Bacon, S. P. and Grantham, D. W. 1992. Fringe effects in modulation masking. *J.Acoust.Soc.Am.*, 91:3451–3455.

Bacon, S. P. and Konrad, D. L. 1993. Modulation detection interference under conditions favoring within- or across-channel processing. *J.Acoust.Soc.Am.*, 93:1012–1022.

Bacon, S. P. and Moore, B. C. J. 1993. Modulation detection interference: some spectral effects. *J.Acoust.Soc.Am.*, 93:3442–3453.

Bacon, S. P., Moore, B. C. J., Shailer, M. J., and Jorasz, U. 1995. Effects of combining maskers in modulation detection interference. *J.Acoust.Soc.Am.*, 97:1847–1853.

Bacon, S. P. and Opie, J. M. 1994. Monotic and dichotic modulation detection interference in practiced and unpracticed subjects. *J.Acoust.Soc.Am.*, 95:2637–2641.

Bacon, S. P. and Opie, J. M. 2002. Modulation detection interference in listeners with normal and impaired hearing. *J.Speech Lang Hear.Res.*. 45:392–402.

Bacon, S. P., Opie, J. M., and Montoya, D. Y. 1998. The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *J.Speech Lang Hear.Res.*, 41:549–563.

Bacon, S. P. and Viemeister, N. F. 1985. Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners. *Audiology*. 24:117–134.

Berg, B. G. 1996. On the relation between comodulation masking release and temporal modulation transfer functions. *J.Acoust.Soc.Am.*, 100:1013–1023.

Bregman, A. S. 1994. *Auditory Scene Analysis - The perceptual Organization of Sound.* The MIT Press, Cambridge Massachusets.

Bregman, A. S., Abramson, J., Doehring, P., and Darwin, C. J. 1985. Spectral integration based on common amplitude modulation. *Percept.Psychophys.*, 37:483–493.

Bregman, A. S., Levitan, R., and Liao, C. 1990. Fusion of auditory components: effects of the frequency of amplitude modulation. *Percept.Psychophys.*, 47:68–73.

Brüel & Kjær. 1989. printed by brüel & kjær pro audio. *Compact Disc.*, pages 1–54.

Buus, S. and Florentine, M. 2002. Growth of loudness in listeners with cochlear hearing losses: recruitment reconsidered. *J.Assoc.Res.Otolaryngol.*, 3:120–139.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Kotby, M. N., Nasser, N. H. A., El Kholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G. I., Westermann, S., and Ludvigsen, C. 1994. An international comparison of long-term average speech spectra. *J. Acoust. Soc. Am.*, 96: 2108–2120.

Coombs, S. and Fay, R. R. 1985. Adaptation effects on amplitude modulation detection: behavioral and neurophysiological assessment in the goldfish auditory system. *Hear.Res.*, 19:57–71.

Cronbach, L. J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16:297–333.

Dau, T. 1996. *Modeling auditory processing of amplitude modulation.* Phd-thesis, der Carl von Ossietzky Universitt Oldenburg.

Dau, T., Kollmeier, B., and Kohlrausch, A. 1997a. Modeling auditory signal processing of amplitude modulation. II. spectral and temporal integration. *J.Acoust.Soc.Am.*, 102: 2906–2919.

Dau, T., Kollmeier, B., and Kohlrausch, A. 1997b. Modeling auditory signal processing of amplitude modulation. I. detection and masking with narrow-band carriers. *J.Acoust.Soc.Am.*, 102:2892–2905.

Dau, T., Verhey, J., and Kohlrausch, A. 1999. Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers. *J.Acoust.Soc.Am.*, 106: 2752–2760.

Dorman, M. F., Loizou, P. C., and Rainey, D. 1997. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *J.Acoust.Soc.Am.*, 102:2403–2411.

Dreschler, W. A. 1983. *Relations between psychophysical data and speech perception for hearing-impaired subjects.* PhD thesis, Rodopi Amsterdam.

Dreschler, W. A. and Leeuw, A. R. 1990. Speech reception in reverberation related to temporal resolution. *J.Speech Hear.Res.*, 33:181–187.

Dreschler, W. A., Verschuure, J., Ludvigsen, C., and Westermann, S. 2001. Icra noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. international collegium for rehabilitative audiology. *Audiology*, 40:148–157.

Drullman, R. 1995. Temporal envelope and fine structure cues for speech intelligibility. *J.Acoust.Soc.Am.*, 97:585–592.

Drullman, R., Festen, J. M., and Plomp, R. 1994a. Effect of reducing slow temporal modulations on speech reception. *J.Acoust.Soc.Am.*, 95:2670–2680.

Drullman, R., Festen, J. M., and Plomp, R. 1994b. Effect of temporal envelope smearing on speech reception. *J.Acoust.Soc.Am.*, 95:1053–1064.

Duifhuis, H. 1973. Consequences of peripheral frequency selectivity for non-simultaneous masking. *J.Acoust.Soc.Am.*, 54:1471–1488.

Duquesnoy, A. J. and Plomp, R. 1980. Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis. *J. Acoust. Soc. Am.*, 68:537–544.

Eddins, D. A. 1993. Amplitude modulation detection of narrow-band noise: Effects of absolute bandwidth and frequency region. *J.Acoust.Soc.Am.*, 93:470–479.

Eddins, D. A. 1999. Amplitude-modulation detection at low- and high-audio frequencies. *J.Acoust.Soc.Am.*, 105:829–837.

Eddins, D. A., Hall, III, J. W., and Grose, J. H. 1992. The detection of temporal gaps as a function of frequency region and absolute noise bandwidth. *J.Acoust.Soc.Am.*, 91: 1069–1077.

Eddins, D. A. and Wright, B. A. 1994. Comodulation masking release for single and multiple rates of envelope fluctuation. *J.Acoust.Soc.Am.*, 96:3432–3442.

Eggermont, J. J. 2002. Temporal modulation transfer functions in cat primary auditory cortex: separating stimulus effects from neural mechanisms. *J.Neurophysiol.*, 87:305–321.

Fantini, D. A. 1991. The processing of envelope information in comodulation masking release (CMR) and envelope discrimination. *J.Acoust.Soc.Am.*, 90:1876–1888.

Festen, J. M. 1987. *The psychophysics of Speech Reception, edited by Schouten,M. E. H.*, chapter Speech-reception threshold in a fluctuating background sound and its possible relation to temporal auditory resolution, pages 461–466. Martinus Nijhoff Publishers.

Festen, J. M. 1993. Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice. *J.Acoust.Soc.Am.*, 94: 1295–1300.

Festen, J. M. and Plomp, R. 1983. Relations between auditory functions in impaired hearing. *J.Acoust.Soc.Am.*, 73:652–662.

Festen, J. M. and Plomp, R. 1990. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J.Acoust.Soc.Am.*, 88: 1725–1736.

Fletcher, H. 1940. Auditory patterns. *Rev.Mod.Phys.*, 12:47–65.

Formby, C. and Muir, K. 1988. Modulation and gap detection for broadband and filtered noise signals. *J.Acoust.Soc.Am.*, 84:545–550.

Forrest, T. G. and Green, D. M. 1987. Detection of partly filled gaps in noise and the temporal modulation transfer function. *J.Acoust.Soc.Am.*, 82:1933–1943.

Franck, B. A. M., Lyzenga, J., and Dreschler, W. A. 2004. Optimal combinations of three speech-processing strategies for hearing-impaired subjects. *to be submitted.*

Frisbie, D. A. 1988. Reliability of scores from teacher-made tests. *Educational Measurements: Issues and practice, vol. 7*, pages 53–63.

Fu, Q. and Shannon, R. V. 1998. Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners. *J.Acoust.Soc.Am.*, 104:2570-2577.

Gardner, R. B. and Wilson, J. P. 1979. Evidence for direction-specific channels in the processing of frequency modulation. *J.Acoust.Soc.Am.*, 66:704–709.

Glasberg, B. R. and Moore, B. C. J. 1989. Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. *Scand.Audiol.Suppl*, 32:1–25.

Glasberg, B. R. and Moore, B. C. J. 1990. Derivation of auditory filter shapes from notched-noise data. *Hear.Res.*, 47:103–138.

Glasberg, B. R., Moore, B. C. J., and Bacon, S. P. 1987. Gap detection and masking in hearing-impaired and normal-hearing subjects. *J.Acoust.Soc.Am.*, 81:1546–1556.

Grantham, D. W. and Bacon, S. P. 1991. Binaural modulation masking. *J.Acoust.Soc.Am.*, 83:1340.

Greenberg, S., Arai, T., and Silipo, R. 1998. Speech intelligibility derived from exceedingly sparse spectral information. *Proceedings of the international conference on spoken language processing. Sydney.* pages 1–4.

Greenwood, D. D. and Joris, P. X. 1996. Mechanical and "temporal" filtering as codeterminants of the response by cat primary fibers to amplitude-modulated signals. *J.Acoust.Soc.Am.*, 99:1029–1039.

Grose, J. H. and Hall, III, J. W. 1992. Comodulation masking release for speech stimuli. *J.Acoust.Soc.Am.*, 91:1042–1050.

Grose, J. H. and Hall, III, J. W. 1994. Modulation detection interference (MDI) in listeners with cochlear hearing loss. *J.Speech Hear.Res.*, 37:680–686.

Grose, J. H. and Hall, III, J. W. 1996. Cochlear hearing loss and the processing of modulation: effects of temporal asynchrony. *J.Acoust.Soc.Am.*, 100:519–527.

Gustafsson, H. A. and Arlinger, S. D. 1994. Masking of speech by amplitude modulated noise. *J.Acoust.Soc.Am.*, 95:518–529.

Hall, III, J. W. and Grose, J. H. 1988. Comodulation masking release: evidence for multiple cues. *J.Acoust.Soc.Am.*, 84:1669–1675.

Hall, III, J. W. and Grose, J. H. 1989. Spectrotemporal analysis and cochlear hearing impairment: effects of frequency selectivity, temporal resolution, signal frequency, and rate of modulation. *J.Acoust.Soc.Am.*, 85:2550–2562.

Hall, III, J. W. and Grose, J. H. 1991. Some effects of auditory grouping factors on modulation detection interference (MDI). *J.Acoust.Soc.Am.*, 90:3028–3035.

Hall, III, J. W. and Grose, J. H. 1992. Masking release for gap detection. *Philos.Trans.R.Soc.Lond B Biol.Sci.*, 336:331–337.

Hall, III, J. W. and Grose, J. H. 1994. Signal detection in complex comodulated backgrounds by normal-hearing and cochlear-impaired listeners. *J.Acoust.Soc.Am.*, 95:435–443.

Hall, III, J. W. and Grose, J. H. 1995. Amplitude discrimination in masking release paradigms. *J.Acoust.Soc.Am.*, 98:847–852.

Hall, III, J. W., Grose, J. H., Buss, E., and Hatch, D. R. 1998. Temporal analysis and stimulus fluctuation in listeners with normal and impaired hearing. *J.Speech Lang Hear.Res.*, 41:340–354.

Hall, III, J. W., Grose, J. H., and Haggard, M. P. 1990. Effects of flanking band proximity, number, and modulation pattern on comodulation masking release. *J.Acoust.Soc.Am.*, 87:269–283.

Hall, III, J. W., Haggard, M. P., and Fernandes, M. A. 1984. Detection in noise by spectro-temporal pattern analysis. *J.Acoust.Soc.Am.*, 76:50–56.

Hartmann, W. M. and Pumplin, J. 1988. Noise power fluctuations and the masking of sine signals. *J.Acoust.Soc.Am.*, 83:2277-2289.

Hearing Journal. 1998. Audio cd accompnying articles in the hearing journal. *Hearing Journal*, pages 1–37.

Hellbrück, J. and Moser, L. M. 1986. Psychophysical measurements of inner ear processes with category scaling of loudness. *ORL J.Otorhinolaryngol.Relat Spec.*, 48:135–136.

Horst, J. W. 1987. Frequency discrimination of complex signals, frequency selectivity, and speech perception in hearing-impaired subjects. *J.Acoust.Soc.Am.*, 82:874–885.

Houtgast, T. 1989. Frequency selectivity in amplitude-modulation detection. *J.Acoust.Soc.Am.*, 85:1676–1680.

Houtgast, T. and Steeneken, H. J. M. 1973. The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acustica*, 28:66–73.

Howard-Jones, P. A. and Rosen, S. 1993a. The perception of speech in fluctuating noise. *J.Acoust.Soc.Am.*, 78:258–272.

Howard-Jones, P. A. and Rosen, S. 1993b. Uncomodulated glimpsing in "checkerboard" noise. *J.Acoust.Soc.Am.*, 93:2915–2922.

Javel, E. 1996. Long-term adaptation in cat auditory-nerve fiber responses. *J.Acoust.Soc.Am.*, 99:1040–1052.

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

Joris, P. X. and Yin, T. C. 1992. Responses to amplitude-modulated tones in the auditory nerve of the cat. *J.Acoust.Soc.Am.*, 91:215–232.

Kay, R. H. and Matthews, D. R. 1972. On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones. *J.Physiol*, 225:657–677.

Keidser, G. 1995. Long-term spectra of a range of real-life noisy environments. *Austr.Aud.*, 17:39–46.

Kleinbaum, D. G., Kupper, L. K., and Muller, K. E. 1987. *Applied Regression Analysis and Other Multivariable Methods*, chapter Variable Reduction and Factor Analysis. PWS-KENT.

Köbler, S. and Leijon, A. 1999. Noise analysis of real-life listening situations for maximal speech audibility in hearing aid fitting. *Scand.Audiol.*, 28:179–189.

Kohlrausch, A. 1993. Comment on "temporal modulation transfer functions in patients with cochlear implants" [J. Acoust. Soc. Am. 91,2156-2164 (1992)]. *J.Acoust.Soc.Am.*, 93:1649–1650.

Kohlrausch, A., Fassel, R., and Dau, T. 2000. The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. *J.Acoust.Soc.Am.*, 108:723-734.

Kuwada, S., Anderson, J. S., Batra, R., Fitzpatrick, D. C., Teissier, N., and D'Angelo, W. R. 2002. Sources of the scalp-recorded amplitude-modulation following response. *J.Am.Acad.Audiol.*, 13:188-204.

Kwon, B. J. 2002. Comodulation masking release in consonant recognition. *J.Acoust.Soc.Am.*, 112:634-641.

Kwon, B. J. and Turner, C. W. 2001. Consonant identification under maskers with sinusoidal modulation: masking release or modulation interference ? *J.Acoust.Soc.Am.*, 110:1130–1140.

Larsby, B. and Arlinger, S. 1994. Speech recognition and just-follow-conversation tasks for normal-hearing and hearing-impaired listeners with different maskers. *Audiology*, 33: 165-176.

Lawson, J. L. and Uhlenbeck, G. E. 1950. *Threshold Signals, volume 24 of Radiation Laboratory Series*. McGraw Hill, New York.

Lee, J. and Bacon, S. P. 1997. Amplitude modulation depth discrimination of a sinusoidal carrier: effect of stimulus duration. *J.Acoust.Soc.Am.*, 101:3688–3693.

Leeuw, A. R. and Dreschler, W. A. 1994. Frequency-resolution measurements with notched noises for clinical purposes. *Ear Hear.*, 15:240–255.

Levitt, H. 1971. Transformed up-down methods in psychoacoustics. *J.Acoust.Soc.Am.*, 49: 467–477.

Lippman, R. 1996. Accurate consonant perception without mid-frequency speech energy. *IEEE Trans Speech Audio Proc*, 4:66-69.

Lorenzi, C., Micheyl, C., Berthommier, F., and Portalier, S. 1997. Modulation masking in listeners with sensorineural hearing loss. *J.Speech Lang Hear.Res.*, 40:200–207.

Lyregaard, P. E. 1988. *Hearing aid fitting, theoretical and practical views, edited by Jensen, J. H.*, chapter Pogo and the theory behind it, pages 259-284. Stougaard Jensen.

Mäkelä, J. P., Hari, R., and Linnankivi, A. 1987. Different analysis of frequency and amplitude modulations of a continuous tone in the human auditory cortex: a neuromagnetic study. *Hear.Res.*, 27:257–264.

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

Mendoza, L., Hall, III, J. W., and Grose, J. H. 1995a. Modulation detection interference using random and sinusoidal amplitude modulation. *J.Acoust.Soc.Am.*, 97:2487–2492.

Mendoza, L., Hall, III, J. W., and Grose, J. H. 1995b. Within- and across-channel processes in modulation detection interference. *J.Acoust.Soc.Am.*, 97:3072–3079.

Miller, G. A. and Licklider, J. C. 1950. The intelligibility of interrupted speech. *J.Acoust.Soc.Am.*, 22:167–173.

Millman, R. E., Lorenzi, C., Apoux, F., Fullgrabe, C., Green, G. G., and Bacon, S. P. 2002. Effect of duration on amplitude-modulation masking. *J.Acoust.Soc.Am.*, 111:2551–2554.

Møller, A. R. 1974. Responses of units in the cochlear nucleus to sinusoidally amplitude-modulated tones. *Exp.Neurol.*, 45:105–117.

Moore, B. C. and Emmerich, D. S. 1990. Monaural envelope correlation perception, revisited: effects of bandwidth, frequency separation, duration, and relative level of the noise bands. *J.Acoust.Soc.Am.*, 87:2628–2633.

Moore, B. C., Shailer, M. J., Hall, III, J. W., and Schooneveldt, G. P. 1993. Comodulation masking release in subjects with unilateral and bilateral hearing impairment. *J.Acoust.Soc.Am.*, 93:435–451.

Moore, B. C. J. 1992. Across-channel processes in auditory masking. *J.Acoust.Soc.Jpn.*, 13:25–37.

Moore, B. C. J. and Glasberg, B. R. 1987. Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hear.Res.*, 28:209–225.

Moore, B. C. J., Glasberg, B. R., Donaldson, E., McPherson, T., and Plack, C. J. 1989. Detection of temporal gaps in sinusoids by normally hearing and hearing-impaired subjects. *J.Acoust.Soc.Am.*, 85:1266–1275.

Moore, B. C. J., Glasberg, B. R., Gaunt, T., and Child, T. 1991. Across-channel masking of changes in modulation depth for amplitude- and frequency-modulated signals. *Q.J.Exp.Psychol.A*, 43:327–347.

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. and Jorasz, U. 1996. Modulation discrimination interference and comodulation masking release as a function of the number and spectral placement of narrow-band noise modulators. *J.Acoust.Soc.Am.*, 100:2373–2381.

Moore, B. C. J. and Shailer, M. J. 1992. Modulation discrimination interference and auditory grouping. *Philos.Trans.R.Soc.Lond B Biol.Sci.*, 336:339–346.

Moore, B. C. J. and Shailer, M. J. 1994. Effects of harmonicity, modulator phase, and number of masker components on modulation discrimination interference. *J.Acoust.Soc.Am.*, 95:3555–3560.

Moore, B. C. J., Vickers, D. A. Baer, T., and Launer, S. 1999. Factors affecting the loudness of modulated sounds. *J.Acoust.Soc.Am.*, 105:2757–2772.

Moore, B. C. J., Wojtczak, M., and Vickers, D. A. 1996. Effect of loudness recruitment on the perception of amplitude modulation. *J.Acoust.Soc.Am.*, 100:481–489.

Müller-Preuss, P., Flachskamm, C., and Bieser, A. 1994. Neural encoding of amplitude modulation within the auditory midbrain of squirrel monkeys. *Hear.Res.*, 80:197–208.

Neuert, V., Pressnitzer, D., Patterson, R. D., and Winter, I. M. 2001. The responses of single units in the inferior colliculus of the guinea pig to damped and ramped sinusoids. *Hear.Res.*, 159:36–52.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. 2000. Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. II. hearing-impaired listeners. *J.Acoust.Soc.Am.*, 107:1685–1696.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. 2001. Relations between intelligibility of narrow-band speech and auditory functions, both in the 1-kHz frequency region. *J.Acoust.Soc.Am.*, 109:1197–1212.

Oticon. 1993. printed by. the sensation of multifocus. *Compact Disc.*, 1:1–22.

Oxenham, A. J. 1997. Increment and decrement detection in sinusoids as a measure of temporal resolution. *J.Acoust.Soc.Am.*, 102:1779–1790.

Oxenham, A. J. and Dau, T. 2001. Modulation detection interference: effects of concurrent and sequential streaming. *J.Acoust.Soc.Am.*, 110:402–408.

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

Ozimek, E. and Sek, A. 1988. Am diiferelnce limens for noise bands. *Acustica*. 66:153–160.

Pavlovic, C. V. 1984. Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment. *J.Acoust.Soc.Am.*, 75:1253–1258.

Pavlovic, C. V. 1987. Derivation of primary parameters and procedures for use in speech intelligibility predictions. *J.Acoust.Soc.Am.*, 82:413–422.

Peters, R. W. and Hall, III, J. W. 1994. Comodulation masking release for elderly listeners with relatively normal audiograms. *J.Acoust.Soc.Am.*, 96:2674–2682.

Phonak. 1993. printed by. phonak environments. *Compact Disc.*, 1:1–62.

Pinheiro, J. C. and Bates, D. M. 2000. *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag, New York.

Plack, C. J. and Oxenham, A. J. 1997. A behavioral measure of basilar-membrane non-linearity in listeners with normal and impaired hearing. *J.Acoust.Soc.Am.*, 101:3666–3675.

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

Plack, C. J., Oxenham, A. J., and Drga, V. 2002. Linear and nonlinear processes in temporal masking. *Acta Acustica*, 88:348–358.

Plasmans, C. J., Koopman, J., and Dreschler, W. A. 2003. The effects of frequency and bandwidth on loudness perception, measured by magnitude production and categorical loudness scaling. *Unpublished data.*

Plomp, R. 1964. Rate of decay of auditory sensation. *J.Acoust.Soc.Am*, 36:277–282.

Plomp, R. 1986. A signal-to-noise ratio model for the speech-reception threshold of the hearing-impaired. *J.Speech Hear.Res.*, 29:146–154.

Plomp, R., Houtgast, T., and Steeneken, H. J. M. 1984. *Limited in Perception, edited by van Doorn, A. J. and van de Grind, A. and Koenderink, J.*, chapter 5 The modulation transfer function in audition. VNU Science Press, Utrecht.

Plomp, R. and Mimpen, A. M. 1979. Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, 18:43–52.

Rees. A.. Green. G. G.. and Kay, R. H. 1986. Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man. *Hear.Res.*. 23:123–133.

Rees. A. and Moller, A. R. 1983. Responses of neurons in the inferior colliculus of the rat to am and fm tones. *Hear.Res.*. 10:301–330.

Regan. D. and Tansley. B. W. 1979. Selective adaptation to frequency-modulated tones: evidence for an information-processing channel selectively sensitive to frequency changes. *J.Acoust.Soc.Am.*. 65:1249–1257.

Richards. V. M. 1992. The detectability of a tone added to narrow bands of equal-energy noise. *J.Acoust.Soc.Am.*. 91:3424–3435.

Richards. V. M.. Buss. E.. and Tian. L. 1997. Effects of modulator phase for comodulation masking release and modulation detection interference. *J.Acoust.Soc.Am.*. 102:468–476.

Robles. L. Ruggero. M. A. and Rich. N. C. 1986. Basilar membrane mechanics at the base of the chinchilla cochlea. I. input-output functions. tuning curves. and response phases. *J.Acoust.Soc.Am.* 80:1364–1374.

Rosen, S. 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 336(1278):367–373.

Schlauch. R. S., DiGiovanni, J. J., and Ries, D. T. 1998. Basilar membrane nonlinearity and loudness. *J.Acoust.Soc.Am*, 103:2010–2020.

Schmidt. A. M. G. 1954. *Jip en Janneke deel 1.* Querido kinderboeken.

Schreiner. C. E. and Urbas, J. V. 1986. Representation of amplitude modulation in the auditory cortex of the cat. I. the anterior auditory field (AAF). *Hear.Res.*, 21:227–241.

Shailer. M. J. and Moore. B. C. J. 1983. Gap detection as a function of frequency, bandwidth. and level. *J.Acoust.Soc.Am*, 74:467–473.

Shailer. M. J. and Moore. B. C. J. 1985. Detection of temporal gaps in bandlimited noise: effects of variations in bandwidth and signal-to-masker ratio. *J.Acoust.Soc.Am*, 77:635–639.

Shailer. M. J. and Moore. B. C. J. 1993. Effects of modulation rate and rate of envelope change on modulation discrimination interference. *J.Acoust.Soc.Am.*. 94:3138–3143.

Shannon, R. V. 1990. Forward masking in patients with cochlear implants. *J.Acoust.Soc.Am.*, 88:741–744.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. 1995. Speech recognition with primarily temporal cues. *Science*. 270:303–304.

Siemens. 1992. printed by. anpassung und test von hörprogrammen. *Compact Disc.*, 1: 1–56.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. 2002. Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416:87–90.

Sound Ideas. 1988. Sound ideas; sound effects library. *Compact Disc.*, 1–14:1–17.

Steinberg, J. C. and Gardner, M. B. 1937. The dependence of hearing-impairment on sound intensity. *J.Acoust.Soc.Am.*, 9:11–23.

Strickland, A. E. and Viemeister, N. F. 1996. Cues for discrimination of envelopes. *J.Acoust.Soc.Am.*, 99:3638–3646.

Takahashi, G. A. and Bacon, S. P. 1992. Modulation detection, modulation masking, and speech understanding in noise in the elderly. *J.Speech Hear.Res.*, 35:1410–1421.

Tansley, B. W. and Regan, D. 1979. Separate auditory channels for unidirectional frequency modulation and unidirectional amplitude modulation. *Sens.Processes*, 3:132–140.

Tansley, B. W. and Suffield, J. B. 1983. Time course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation. *J.Acoust.Soc.Am.*, 74:765–775.

TNO. 1990. printed by. speech research institute, rsre - noise-rom-0 - nato: Ac243/(panel 3)/rsg-10; esprit: Project no. 2589-sam. *Compact Disc.*, 1:1–19.

Trine, T. D. 1995. *Speech recognition in modulated noise and temporal resolution: effects of listening bandwidth.* PhD thesis. University of Minnesota.

van de Par, S. and Kohlrausch, A. 1998. Comparison of monaural (CMR) and binaural (BMLD) masking release. *J.Acoust.Soc.Am.*, 103:1573–1579.

van der Horst, R., Leeuw, A. R., and Dreschler, W. A. 1999. Importance of temporal-envelope cues in consonant recognition. *J.Acoust.Soc.Am.*, 105:1801–1809.

van Schijndel, N. H. 2000. *Wavelet coding of sound as a tool for studying the auditory system.* PhD thesis. Vrije Universiteit Amsterdam.

Verhey, J. L., Dau, T., and Kollmeier, B. 1999. Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation filter bank. *J.Acoust.Soc.Am.*, 106:2733–2745.

Verschuure, J., Benning, F. J., van Cappellen, M., Dreschler, W. A., and Boermans, P. P. 1998. Speech intelligibility in noise with fast compression hearing aids. *Audiology*, 37: 127–150.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. 2000. Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J.Acoust.Soc.Am.*, 107:1671–1684.

Versfeld, N. J. and Dreschler, W. A. 2002. The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners. *J.Acoust.Soc.Am.*, 111:401–408.

Viemeister, N. F. 1979. Temporal modulation transfer functions based upon modulation thresholds. *J.Acoust.Soc.Am.*, 66:1364–1380.

Viemeister, N. F. and Wakefield, G. H. 1991. Temporal integration and multiple looks. *J.Acoust.Soc.Am.*, 90:858–865.

von Fleischer, H. 1980. Subjektive größe von unterschieden im amplituden-modulationsgrad von sinustönen. *Acustica*, 46:31–38.

Wakefield, G. H. and Viemeister, N. F. 1990. Discrimination of modulation depth of sinusoidal amplitude modulation (sam) noise. *J.Acoust.Soc.Am.*, 88:1367–1373.

Widex. 1995. printed by. real-life environment sound examples. *Compact Disc.*, 1:1 19.

Wojtczak, M. 1996. *Psychoacoustics, speech and hearing aids, edited by Kollmeier, B.*, chapter Perception of intensity and frequency modulation in people with normal and impaired hearing, pages 35–38. World Scientific Publishing Co. Pte. Ltd.

Wojtczak, M. and Viemeister, N. F. 2003. Suprathreshold effects of adaptation produced by amplitude modulation. *J.Acoust.Soc.Am.*, 114:991–997.

Yost, W. A. and Sheft, S. 1989. Across-critical-band processing of amplitude-modulated tones. *J.Acoust.Soc.Am.*, 85:848–857.

Yost, W. A. and Sheft, S. 1990. A comparison among three measures of cross-spectral processing of amplitude modulation with tonal signals. *J.Acoust.Soc.Am.*, 87:897–900.

Yost, W. A. and Sheft, S. 1994. Modulation detection interference: across-frequency processing and auditory grouping. *Hear.Res.*, 79:48–58.

Yost, W. A., Sheft, S., and Opie, J. 1989. Modulation interference in detection and discrimination of amplitude modulation. *J.Acoust.Soc.Am.*, 86:2138–2147.

Zwicker, E. 1961. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *J.Acoust.Soc.Am.*, 33:248.

Zwicker, E., Flottorp, G., and Stevens, S. S. 1957. Critical bandwidth in loudness summation. *J.Acoust.Soc.Am*, 29:548–557.

Zwislocki, J. J. 1960. Theory of temporal auditory summation. *J.Acoust.Soc.Am.*, 32: 1046–1060.

Stellingen

Behorend bij het proefschrift van Jan Koopman

1. Dat het spraakverstaan voor slechthorenden minder verbetert door het aanbrengen van temporele gaten in de ruis dan voor normaalhorenden, wordt mede veroorzaakt door een afname in de gevoeligheid voor amplitude modulaties. (*dit proefschrift*)

2. Het perceptief groeperen van verschillende geluiden (perceptual grouping) is n van de belangrijkste oorzaken dat modulatie detectie interferentie (MDI) optreedt. Maar effecten zoals adaptatie kunnen daarbij zeker niet buiten beschouwing gelaten worden. (*dit proefschrift hoofdstuk 9*)

3. Fysisch gezien lijkt het nemen van vaste stappen in dB's een logische keuze bij het adaptief testen van de gevoeligheid voor amplitude modulaties. Vanuit de perceptie gezien, is het echter raadzaam om de stapgrootte af te laten hangen van de modulatiediepte waarmee gemeten wordt. (*dit proefschrift hoofdstukken 6 en 10*)

4. Noordhoek (2000) onderscheidt twee groepen van slechthorendheid; zij die problemen ervaren met de temporele resolutie en zij die problemen hebben met de spectrale resolutie. Kortweg, ook wel de f-groep en de t-groep genoemd. Deze studie laat zien dat er misschien een derde groep bestaat; zij die problemen ervaren met het waarnemen van de omhullende, ofwel de m-groep. (*dit proefschrift hoofdstukken 8 en 9*)

5. Omdat het modelleren in de psychofysica in het algemeen gebeurt aan de hand van veel verschillende condities voor een beperkt aantal proefpersonen, zouden gemengde modellen (linear mixed effects models) als statistisch gereedschap veel meer gebruikt moeten worden.

6. Gehoorbeschermers worden vooral geselecteerd op basis van de verzwakking die geboden wordt. Het zijn echter vooral de effecten op spraakverstaan en lokalisatie die de bruikbaarheid in de praktijk bepalen. Deze dienen dan ook binnen de specificaties te worden opgenomen.

7. Als iedereen kennis zou hebben van de logaritmische schaal, zou het een stuk gemakkelijker worden om de noodzaak van gehoorbeschermers duidelijk te maken.

8. De mogelijkheden van de statistiek zijn ongekend. Het stoeien met de statistiek kan dan ook tot onvoorstelbare resultaten leiden (*zie omslag*).

9. Het oor is als een computer. Al experimenterend merk je pas wat de echte mogelijkheden zijn.

10. Het schrijven van een proefschrift is als het fietsen op een alpenreus. In het begin kijk je er naar uit, halverwege zie je er tegen op, aan het einde heb je het zwaar, op de top ben je blij dat het erop zit.

This thesis addresses a highly relevant aspect of auditory perception: the role of amplitude modulations. The work brings together: the role of modulations in signal classification; the perception of sinusoidally amplitude modulated signals; and the importance of modulation perception for speech intelligibility. The sensitivity to amplitude modulated signals is not only adversely affected when a modulated masker and target are presented simultaneously, but also by pre-exposure to a modulated masker, especially for hearing-impaired subjects. This study shows that relatively short pre-exposure times reduce the sensitivity to amplitude modulated signals for hearing-impaired subjects. The research described in this thesis investigates whether this may contribute to the reduced ability of hearing-impaired subjects to use fluctuations in a background noise to improve speech intelligibility. The findings presented in this thesis suggest that a reduced sensitivity to sinusoidally amplitude modulated signals following pre-exposure to sinusoidally amplitude modulation may indeed affect speech intelligibility in a fluctuating noise. The consequences of this study may be of interest to those developing innovative speech tests or hearing aid algorithms.