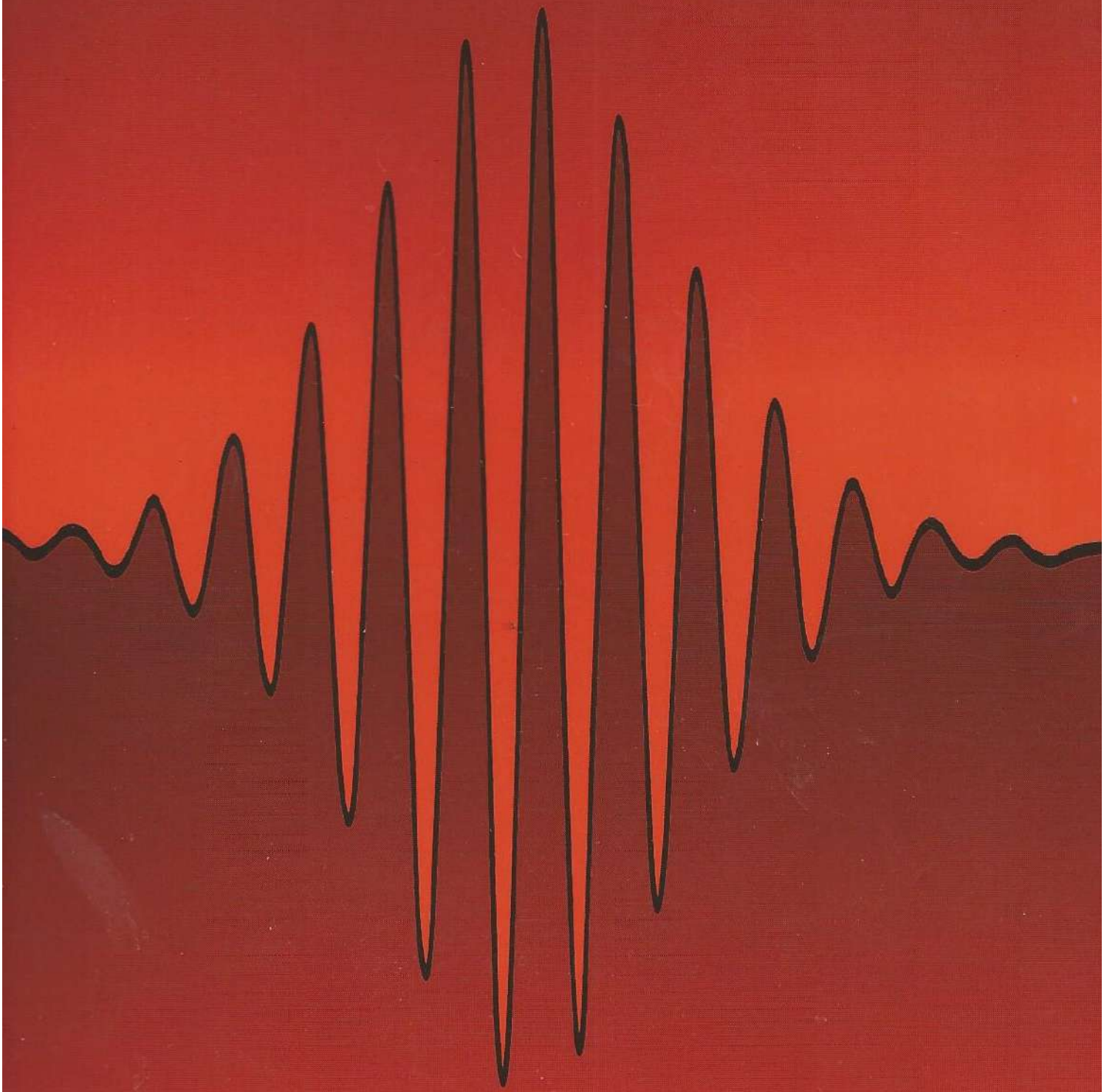
Wavelet coding of sound as a tool for studying the auditory system



Nicolle H. van Schijndel

Wavelet coding of sound as a tool for studying the auditory system

VRIJE UNIVERSITEIT

WAVELET CODING OF SOUND AS A TOOL FOR STUDYING THE AUDITORY SYSTEM

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan
de Vrije Universiteit te Amsterdam,
op gezag van de rector magnificus
prof.dr. T. Sminia,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der geneeskunde
op woensdag 26 april 2000 om 13.45 uur
in het hoofdgebouw van de universiteit,
De Boelelaan 1105

door

Nicolle Hanneke van Schijndel

geboren te Heeswijk-Dinther

Promotor: prof.dr.ir. T. Houtgast

Copromotor: dr.ir. J.M. Festen

Aan mama en papa

This research project was supported by
the Netherlands Organization for Scientific Research (NWO).

Printing: Ponsen & Looijen

ISBN: 906464697x

CONTENTS

1 General introduction

In this thesis wavelet coding of sound is used as a tool to study the auditory system. The type of wavelet coding used is important. Parameters of the wavelet analysis should be tailored to the auditory system. Therefore, in the first part of this thesis, a perceptually relevant wavelet analysis and synthesis scheme is developed. In the second part, this scheme is used as a front-end signal processing tool for studying which auditory coding deficits impair speech perception in hearing-impaired listeners.

I. WAVELET CODING

Wavelets are "little waves that start and stop" (Strang, 1994). Sounds can be decomposed into wavelets, just as sounds can be decomposed into sines and cosines, as is done in Fourier analysis. A wavelet can be considered to represent a time-frequency window. Each wavelet originates from one prototype analysis function, the mother wavelet. A basis of wavelets is constructed by compression (or stretching) of this prototype function to cover the frequency domain, and by shifting of this prototype function to cover the temporal domain. Wavelet coding will be explained in more detail in Chapter 3 of this thesis, but discussing the mother wavelet briefly here seems useful. In Fourier analysis,

one is bounded to use sines and cosines. However, in wavelet analysis, one can choose among different mother wavelets. The choice of the mother wavelet determines the shape of the corresponding spectro-temporal analysis window, thus determining the temporal and spectral resolution of the wavelet analysis. As a result, a different choice of the mother wavelet will result in a different spectro-temporal representation of sound. Thus, the choice of the mother wavelet is important.

II. WAVELET CODING OF SOUND AS A TOOL FOR STUDYING THE AUDITORY SYSTEM?

In the field of signal analysis, the decomposition of a signal in wavelets is a recent development. When applied to sound, the wavelet approach results in a type of time-frequency representation that matches, to some extent, the properties of sound coding in the auditory system. The spectral resolution of the auditory system is roughly constant on a logarithmic frequency scale. Wavelet analysis uses a logarithmic frequency scale as well. In contrast, in Fourier analysis, spectral resolution is constant on a linear frequency scale. In the past, short-time Fourier analysis was used successfully in the study of the auditory system. Simulating more closely the spectral resolution of the auditory system, wavelet analysis promises to work even better.

III. AIM OF THE THESIS

The aim of this thesis is to investigate how wavelet coding can be used as a tool in psychoacoustics, more specifically, as a tool to study speech perception. Since the type of wavelet coding used is important, first it will be determined which wavelet expansion characterizes auditory spectro-temporal coding as closely as possible. This wavelet coding can be considered a representation of sound that mirrors the properties of auditory

coding. The wavelet coefficients that result from the wavelet coding can be manipulated, introducing specific malformations of the characteristics of sound. Specific artificial distortions introduced in the wavelet coded sounds can be used to study the relevance of certain aspects of auditory coding for speech perception. For instance, the perceptual consequence of a reduced acuity in auditory intensity coding may be simulated by artificially distorting the modulus of wavelet coefficients. In such a way, wavelet coding may provide a powerful tool for studying the relevance of (simulated) changes in certain aspects of auditory coding on sound perception. In this thesis, wavelet coding will be used to study how impaired auditory coding degrades speech perception of hearing-impaired listeners.

IV. DISTORTION-SENSITIVITY MODEL

Roughly speaking, hearing impairment can have two manifestations: (1) reduced audibility, and (2) distortion of perceived sounds (see Plomp, 1978). Because of reduced audibility, sounds have to be presented at higher sound pressure levels than normal to be audible. Because of distortion, sounds that are well above the hearing threshold are subject to some type of distortion process in the ear. The term 'distortion' may recall associations with nonlinear processing. However, the term distortion is not used in this sense in this thesis. Here, it is defined as some kind of deviation from the processing in a normal-hearing listener which is not simply related to an elevated hearing threshold. This auditory distortion leads to so-called suprathreshold auditory deficits that hinder speech processing. The effect of audibility on speech perception is understood well and can be estimated, for example, by the Speech-Intelligibility-Index model (ANSI, 1997). The effects of suprathreshold deficits on speech perception are less clear.

In this study, the effects of distortion of auditory coding on the perception of speech are examined under the so-called distortion-sensitivity model. This model will be explained in more detail in Chapter 4. In the distortion-sensitivity model, performance is measured as a function of some type of artificial distortion. The comparison of the sensitivities to the distortion by normal-hearing and hearing-impaired listeners provides

interesting information, because artificial distortion of cues that are not perceived by the hearing impaired will probably not affect their performance. Thus, if hearing-impaired listeners are less sensitive to the distortion under study than normal-hearing listeners, that type of artificial distortion may relate to the impaired auditory speech coding.

Suprathreshold deficits can express themselves in a distorted processing of intensity, temporal, and spectral information. For example, a deficit that is related to distorted processing of intensity information is impaired loudness perception. Excessive forward masking, i.e., non-simultaneous masking in which a signal is masked by a preceding sound, is an expression of distorted processing of temporal information; excessive upward spread of masking, i.e., simultaneous masking in which a signal is masked by spectral components of lower frequency, is an expression of distorted processing of spectral information. In this thesis, the distortion-sensitivity model will be used to gain insight into the role of a distorted representation of these three types of information for speech perception.

V. OUTLINE

This thesis consists of two parts. In part I (Chapters 2 and 3), the parameters of the wavelet decomposition and recomposition scheme are defined. As explained above, the spectro-temporal shape of the mother wavelet is important for simulating the spectro-temporal resolution of the auditory system. The initial experiment of this thesis was performed to guide the proper choice of the mother wavelet. The results are used to develop a perceptually relevant wavelet analysis and reconstruction method. In part II (Chapters 4 and 5), the wavelet-coding and reconstruction scheme is applied to study the auditory system.

A. Part I: Auditory wavelet coding

The experiments of Chapter 2 aim to characterize the time-frequency window that the normal-hearing ear uses while analyzing sounds. This is done by means of intensity-

discrimination experiments for a specific type of stimulus: a Gaussian-shaped tone pulse. The spectro-temporal shape of this Gaussian tone pulse is varied from narrow-band and long-duration, to broadband and short-duration. Results confirm that auditory spectro-temporal analysis can be modeled well by a wavelet analysis. These results are used to define a mother wavelet that matches the auditory time-frequency window. In Chapter 3, using this mother wavelet, a decomposition and recomposition method is developed, resulting in a perceptually relevant spectro-temporal coding algorithm, i.e., a sound coding method that mirrors, to some extent, the properties of peripheral auditory coding. This method will be used as a front-end signal processing tool for studying the perceptual relevance of (simulated) changes in auditory coding in part II.

B. Part II: The effect of impaired auditory coding on speech perception

Using the perceptually relevant wavelet coding method developed in part I, impaired processing is studied by distortion of the wavelet coefficients between wavelet decomposition and recomposition. When applying this method in a listening experiment, specific manipulations of the wavelet coefficients may be used to simulate specific changes in auditory coding. Thus, the importance of various aspects of auditory coding for speech perception can be investigated.

In Chapter 4, the distortion-sensitivity model is used to study one dimension of auditory coding, i.e., intensity coding. The intensity coding of sound is distorted by random perturbations in the modulus of the wavelet coefficients. Speech intelligibility is measured as a function of this intensity distortion. The aim of this chapter is to investigate if distortion of the intensity information can (partly) explain the reduced speech perception of hearing-impaired listeners.

In Chapter 5, three dimensions of auditory coding, i.e., intensity, time, and frequency, are considered. While Chapter 4 was only concerned with local manipulations in the time-frequency representation of a sound relating to the intensity-coding acuity, in Chapter 5 also spread-of-excitation types of manipulations are studied. This relates to a decrease in the acuity of the spectro-temporal representation of a sound. Temporal and spectral information of sound were distorted by randomly shifting the position of the wavelet coefficients along the time or frequency axis, respectively. The experimental questions are (1) what degrees of distortions are detectable, and (2) how do these

distortions affect speech perception (distortion-sensitivity model). Data with respect to the first question provide information on the acuity of a subject's auditory coding. Data with respect to the second question may give some insight into the practical consequences for speech perception when this acuity is (artificially) reduced. The aim of this chapter is to estimate how impaired auditory coding affects speech perception of hearing-impaired listeners.

In the last chapter of this thesis, a general discussion is given. We will discuss how useful our wavelet coding tool was in revealing the importance of the studied types of information for understanding the suprathreshold deficits underlying poor speech perception by the hearing impaired.

VI. RELATED FIELDS

This thesis aims at a better understanding of the auditory system, especially that of speech perception of hearing-impaired listeners. The typical nature of its approach relates it to several applied topics. For instance, efficient coding and transmission of speech signals is an important area of research. The relatively novel wavelet coding is considered a serious candidate for sparse coding. Our data with respect to detection of coding distortion are related to the minimum number of bits required for speech coding. Our data with respect to the effect of distortion on speech perception may be useful to estimate the reduction in speech intelligibility when speech is sparsely coded.

Insight into how impaired auditory coding affects speech perception may provide important information for the field of speech enhancement. In this field, sound processing algorithms are developed to improve speech intelligibility of hearing-impaired listeners. Insight into what is wrong with auditory coding of hearing-impaired listeners may help to find algorithms that can relieve the speech perception problems caused by degraded auditory coding. The tool developed in this thesis (an auditory-relevant scheme for sound coding and reconstruction) might also be useful for implementation of advanced strategies of signal processing in the time-frequency domain.

REFERENCES

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

Plomp, R. (**1978**). "Auditory handicap of hearing impaired and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.

Strang, G. (**1994**). "Wavelets," American Scientist **82**, 250–255.

2 Intensity discrimination of Gaussian-windowed tones: Indications for the shape of the auditory frequency-time window

The just-noticeable difference in intensity jnd(I) was measured for 1-kHz tones with a Gaussian-shaped envelope as a function of their spectro-temporal shape. The stimuli, with constant energy and a constant product of bandwidth and duration, ranged from a long-duration narrow-band "tone" to a short-duration broadband "click." The jnd(I) was measured in three normal-hearing listeners at sensation levels of 0, 10, 20, and 30 dB in 35 dB(A) SPL pink noise. At intermediate sensation levels, jnd(I) depends on the spectro-temporal shape: at the extreme shapes (tones and clicks), intensity discrimination performance is best, whereas at intermediate shapes the jnd(I) is larger. Similar results are observed at a higher overall sound level, and at a higher carrier frequency. The maximum jnd(I) is observed for stimuli with an effective bandwidth of about 1/3 octave and an effective duration of 4 ms at 1 kHz (1 ms at 4 kHz). A generalized multiple-window model is proposed that assumes that the spectro-temporal domain is partitioned into "internal" auditory frequency-time windows. The model predicts that intensity discrimination thresholds depend upon the number of windows excited by a signal: jnd(I) is largest for stimuli covering one window.

Journal of the Acoustical Society of America 105: 3425-3435, 1999

INTRODUCTION

This study addresses a fundamental psychoacoustical question: how does the auditory system extract spectro-temporal information while processing complex sounds? To obtain information about both the spectral and the temporal structure of a signal, the auditory system performs a frequency-time (f-t) analysis. The result of an f-t analysis is a spectrogram, showing the distribution of signal energy over frequency and time. In the spectrogram, the signal energy at a given point is determined by an integration over a specific frequency-time window. The shape of this f-t analysis window determines which characteristics of the sound are displayed. It is commonly assumed that the peripheral auditory system carries out an f-t analysis with its own specific f-t window. This study seeks to characterize the auditory f-t window.

An f-t analysis window cannot be restricted arbitrarily both in time and in frequency. The minimum area of an f-t window is unity if this area is defined as the product of the effective bandwidth and the effective duration (see Stewart, 1931; Gabor, 1947). The lower bound is attained by the Gaussian function (see Gabor, 1947). Given this restriction, the shape of an f-t analysis window can vary. Different f-t analysis windows will display different details in the f-t representation.

In this study we use a psychophysical approach to gain an insight into the shape of the f-t window underlying auditory sound analysis. Before explaining the experimental paradigm, we will briefly review some current ideas on spectral and temporal resolution in the auditory system and their relation to the auditory f-t window.

In psychoacoustics, the partition of the frequency axis into critical bands is a basic concept. Roughly, the auditory filters have a constant relative bandwidth of 1/3 octave (see, e.g., Scharf and Buus, 1986). This suggests that the spectral width of the auditory f-t window is about 1/3 octave.

In the time domain, however, the picture is less clear. Recall that, for a given bandwidth, the *smallest possible* temporal width is defined by the minimum window area. Thus the temporal width of the auditory f-t window must be at least as long as the minimum temporal width possible, given a specific spectral width. Taking into account psychoacoustical arguments for defining a temporal width, there is no complete

consensus. Plack and Moore (1990) discuss the problem of describing the temporal resolution of the auditory system with a single value. They note that the integration time constant appears to decrease with increasing frequency (see also Gerken *et al.*, 1990; Florentine *et al.*, 1988). This suggests that the temporal width of the auditory f-t window decreases with increasing frequency. Viemeister and Wakefield (1991) are interested in the "resolution-integration" paradox: Models describing temporal resolution use short time constants, whereas models describing the improvement in detection and discrimination with increasing signal duration are based on a process of long-term temporal integration. Their conclusions favor the short time constants (roughly 3 ms for 1-kHz sinusoids). Although temporal integration data (time-intensity trade in detection) can be explained easily by an energy-detector model (single look) with an adjustable time window matched to the signal duration (see, for example, Dai and Wright, 1995), the multiple-look model of Viemeister and Wakefield with short time constants can account for both the data on temporal resolution and temporal integration. In general, temporal resolution experiments suggest that the temporal width of the auditory f-t window is about 3 ms at 1 kHz and smaller at higher frequencies. This is very close to the minimum duration possible if the bandwidth of the auditory f-t window is 1/3 octave.

The present research evaluates the auditory f-t window by assessing just-noticeable differences in intensity for stimuli with different spectro-temporal shapes. By varying the spectro-temporal shape, the number of "internal" (auditory) f-t windows excited by the signal can be varied. Our hypothesis is that this variation affects the just-noticeable difference in intensity. The basis for this hypothesis can be found in the existing models with respect to intensity discrimination.

An important model that describes intensity discrimination performance is the so-called multiband excitation-pattern model (see, e.g., Florentine and Buus, 1981; Durlach *et al.*, 1986; Buus, 1990; Buus and Florentine, 1994). This model operates in the spectral domain. The idea is that the excitation pattern induced by the signal is divided into several spectral bands, and the content of each band is processed individually. Information can be combined across bands to come to an overall percept. Alternatively, psychoacoustical data with respect to temporal mechanisms can be accounted for, at least qualitatively, by the multiple-look model (see Florentine, 1986; Viemeister and Wakefield, 1991). The multiple-look model divides the signal into short-duration segments. As in the multiband model, the information in different segments or "looks"

is considered statistically independent. A combination of different looks will result in more information and therefore in lower thresholds. Conceptually, of course, the multiband model and the multiple-look model are very similar, one operating in the spectral domain and the other in the temporal domain. Durlach *et al.* (1986) note that the frequency channels in their model for discrimination of broadband signals can refer to time intervals as well.

The rationale of the hypothesis of this study is a generalization of these "multi-channel models" (multiband and multiple-look models), which in this paper will be called the "multiple-window model" in the f-t domain. Our hypothesis is the following: the auditory spectro-temporal domain is partitioned into "internal" auditory frequency-time windows. So, the "internal" f-t representation of a signal can be characterized by the number of f-t windows covered by the signal. As in the multiband excitation-pattern model (Florentine and Buus, 1981), the multiple-window model assumes that the discriminability within a window is independent of excitation level (Weber's Law). Thus intensity discriminability for a given signal depends on the number of independent auditory f-t windows covered by the signal: the just-noticeable difference in intensity jnd(I) will be smaller if more windows are involved.

Evaluation of intensity discrimination thresholds for a range of stimuli with well-defined variations in spectro-temporal shape may reveal the shape of the auditory f-t window. By manipulation of the spectro-temporal shape of the stimuli the number of auditory f-t windows covered by the signal can be varied. We are looking for the signal for which the "internal" auditory representation best matches the auditory f-t window. For that purpose we used sinusoids with a Gaussian-shaped temporal envelope. Consequently, it can be shown that the spectral envelope is Gaussian shaped as well. A Gaussian-windowed signal was chosen because of its minimum effective f-t area. Another appealing property is its symmetry in frequency and time. A series of amplitude discrimination experiments was performed for a range of these stimuli in which only one variable was changed, the so-called shape factor, which determines the effective[1] number of periods included under the Gaussian envelope. By varying the shape factor, the representation of the signal in the f-t plane was systematically varied while keeping its

[1]The effective number of periods is defined as the effective duration divided by the period of the carrier frequency of the Gaussian-windowed sinusoid. This is equal to the reciprocal of the shape factor of the signal ($1/\alpha$).

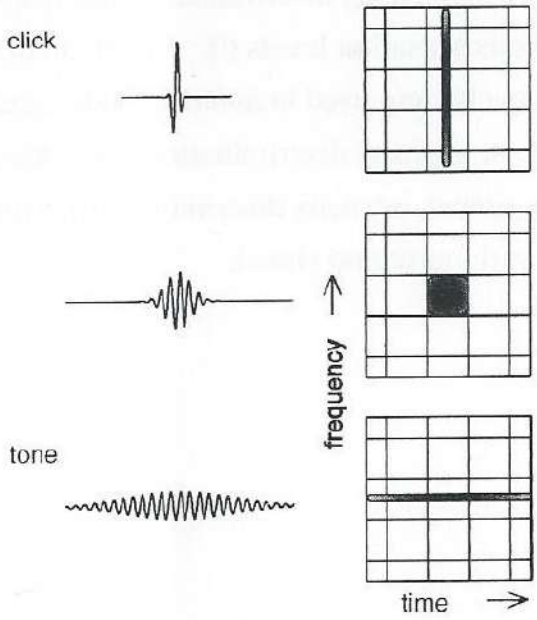


FIG. 2.1. A schematic representation in the f-t plane of three stimuli with different shape factors on a grid of f-t windows.

area constant, ranging from a long-duration narrow-band "tone" to a short-duration broadband "click" (see Fig. 2.1). A tone, corresponding to a small shape factor, will excite many f-t windows along the time axis; a click, corresponding to a large shape factor, will excite many f-t windows along the frequency axis; somewhere between tone and click fewer windows will be excited. Thus the number of auditory f-t windows excited by the signal varies as a function of the shape factor according to an U-shaped curve.

The multiple-window idea states that jnd(I) varies with the number of f-t windows involved in the discrimination task. The more elementary f-t windows which are involved, the smaller the jnd(I) will be. This implies that the signal with the largest jnd(I) covers the minimum number of windows. The shape factor corresponding to this signal will be called the "critical" shape factor. Thus the signal with the critical shape factor is most successful in exciting only the minimum number of f-t windows; the "internal" representation of this signal is most closely related to the elementary f-t window. Within the context of the multiple-window idea, the shape of the f-t representation of that signal best matches the shape of the elementary f-t window in (peripheral) auditory coding. The aim of the experiments is to test the multiple-window hypothesis by examining whether jnd(I) varies with a varying spectro-temporal shape of the stimuli. If jnd(I) varies as a function of shape factor, the critical shape factor gives some insight into the auditory f-t window.

In the first experiment, the relation between intensity discrimination and shape factor was determined for 1-kHz sinusoids at various sensation levels (0,[2] 10, 20, 30 dB SL) in 35-dB(A) SPL pink noise. Low sensation levels were used to avoid spread of excitation as much as possible. In the second experiment, intensity discrimination performance was measured at 4 kHz. Finally, in the third experiment, intensity discrimination performance was measured at a 20-dB higher level for both noise and signal.

I. METHOD

A. Stimuli

The stimuli $s(t)$ consist of Gaussian-windowed tones, defined by

$$s(t)=A\sqrt{\alpha f_0}\sin(2\pi f_0 t+\frac{\pi}{4})\exp(-\pi(\alpha f_0 t)^2) \quad . \tag{2.1}$$

These are sinusoids with carrier frequency f_0 and a gradual onset and offset (see Fig. 2.1). The shape factor α determines the effective number of sinusoidal periods, equal to $1/\alpha$, contained within the Gaussian envelope. If α is small, the number of periods is large (tone). If α is large, the number of periods is small (click). Throughout the experiments the independent variable is the shape factor α (0.0375, 0.075, 0.15, 0.3, 0.6, and 1.2). The effective duration of the Gaussian signal is $\Delta_t=1/(\alpha f_0)$. The effective bandwidth is $\Delta_f=\alpha f_0$.

The amplitude of the signal is defined by $A\sqrt{\alpha f_0}$. The amplitude difference is produced by increasing the amplitude constant A from A_0 to $A_0+\Delta A$. By introducing the phase factor $\pi/4$ the energy of the signal is independent of α and f_0. As a result, the total energy E of the signal is $(\sqrt{2}/4)A^2$, only depending on the amplitude constant A.

---

[2]The intensity discrimination task at 0 dB SL is *not* equal to a detection task, because each interval contains a signal. However, if the reference stimuli are presented at 0 dB SL, the signals are not always audible. When, in a trial, one or two of the stimuli are not audible this is perceived by the subject as a mixture between an amplitude discrimination task and a detection task. Hanna *et al.* (1986) also measured jnd(I) at 0 dB SL.

As already mentioned in the Introduction, in this study the temporal and spectral domain are investigated in combination. Therefore, the set of stimuli in this study consists of stimuli that cover more critical bands along the frequency axis and only one look (about 3 ms at 1 kHz) along the time axis, stimuli that cover only one critical band and more time looks, and stimuli in between. In Table 2.I, the bandwidth, duration, and effective number of periods of the stimuli used in the experiments can be found. The column labeled "# f-t windows" gives the estimated number of f-t windows covered by the 1-kHz tone, assuming a Gaussian auditory f-t window with a shape factor of 0.23, corresponding to a bandwidth of 1/3 octave and a duration of 4 ms at 1 kHz and a duration of 1 ms at 4 kHz.

TABLE 2.I. The effective duration Δ_t, the effective bandwidth Δ_f, the effective number of periods, and the estimated number of f-t windows for stimuli with different shape factor α and carrier frequency f_0 as used in the experiments.

α	f_0	Δ_t	Δ_f	# periods	# f-t windows
0.0375	1000 Hz	27 ms	37.5 Hz	27	7
	4000 Hz	6.7 ms	150 Hz	27	7
0.075	1000 Hz	13 ms	75 Hz	13	3
	4000 Hz	3.3 ms	300 Hz	13	3
0.15	1000 Hz	6.7 ms	150 Hz	6.7	2
	4000 Hz	1.7 ms	600 Hz	6.7	2
0.3	1000 Hz	3.3 ms	300 Hz	3.3	1
	4000 Hz	0.83 ms	1200 Hz	3.3	1
0.6	1000 Hz	1.7 ms	600 Hz	1.7	3
	4000 Hz	0.42 ms	2400 Hz	1.7	3
1.2	1000 Hz	0.83 ms	1200 Hz	0.83	6
	4000 Hz	0.21 ms	4800 Hz	0.83	6

B. Apparatus

Stimuli were generated digitally at a sampling frequency of 40 kHz and were played out over TDT (Tucker Davis Technologies) System II hardware. Because Gaussian-windowed signals do not have compact support,[3] the signals were cutoff at frequencies corresponding to their 60-dB down points. A Wandel und Goltermann RG-1 analog noise generator produced the continuous pink noise. Signals and noise were attenuated (TDT PA4) separately, and subsequently summed (TDT SM3). The stimuli were presented monaurally through Sony MDR-CD999 headphones. Masking noise levels were measured on a Brüel & Kjær type 4152 artificial ear with a flat-plate adapter. The entire experiment was controlled by an IBM PC-compatible computer. Subjects were tested individually in a soundproof room.

C. Procedure

Intensity discrimination performance was measured using an adaptive, three-interval, three-alternative forced-choice paradigm (3I, 3AFC). Each trial consisted of three observation intervals. The time between the onset of the three stimuli was always 500 ms, but the duration of the stimuli differed with different shape factors. Taking into account the cutoff at 60 dB below the top, the total duration of the longest signal was 80 ms. Two intervals contained the reference signal (with amplitude constant A_0) and one interval contained the incremented signal (with amplitude constant $A_0+\Delta A$). The incremented signal occurred randomly in one of the three observation intervals. Each observation interval was marked by a visual display. The onset of the stimuli coincided with the onset of the display. The noise was presented continuously. The subject's task was to indicate the interval that contained the incremented signal by pushing the appropriate button on a PC keyboard. There was no response time limit. Immediately after the response, feedback was provided. After the response, 500 ms elapsed before a following trial started.

---

[3]A function f(t) has compact support if it is zero outside the interval $T_0<t<T_0+\Delta T$.

In obtaining a threshold estimate, the adaptive procedure was started at an increment amplitude ΔA, several steps larger than the anticipated threshold. In the adaptive procedure, the transition from increasing to decreasing difficulty, and vice versa, defined a turnaround. Adaptive thresholds were determined with a one-down/one-up procedure followed by a two-down/one-up procedure after four turnarounds. The steps in the amplitude increment were accomplished by multiplication or division of ΔA by a factor μ (μ<1): $A_{new}=A_0+\mu\Delta A_{old}$ or $A_{new}=A_0+(1/\mu)\Delta A_{old}$, respectively. As a result, the amplitude step in dB gets smaller as the difference in amplitude ΔA between reference signal and incremented signal gets smaller. For the initial steps, μ was 0.66; after four turnarounds, μ was set to 0.8. A run was ended after 24 turnarounds and the geometric mean of the ΔA values of the last 16 turnarounds was used to estimate the threshold ΔA_{jnd}, theoretically equivalent to 70.7% correct (Levitt, 1971). Assuming unbiased responses, the threshold in this paradigm corresponds to a sensitivity d' of about 1.265 (see, e.g., Versfeld *et al.*, 1996). For each subject each condition was repeated six times. The test order of the conditions was balanced according to a Latin square.

Discrimination thresholds were expressed as the just-noticeable difference in intensity, jnd(I) in decibels:

$$jnd(I)=20\log_{10}\frac{A_0+\Delta A_{jnd}}{A_0} \quad , \tag{2.2}$$

where ΔA_{jnd} indicated the amplitude increment yielding 70.7% correct responses.

Beforehand, to set sensation levels for individual subjects and conditions, masked detection thresholds were determined in a similar manner as described above (3 AFC adaptive procedure). Thus, the detection threshold was defined as the threshold at which 70.7% of the stimuli was detected correctly by the listener.

D. Subjects

Three subjects (23–25 years), including the first author, participated in the experiments. All had normal hearing (absolute thresholds better than 15 dB HL at octave frequencies from 125 Hz to 4 kHz and at 6 kHz). Subjects were given practice to stabilize their performance. On the average this took 30 min of practice for five successive days. As a result, practice effects were negligible during the actual experiment.

E. Data analysis

The data analysis was performed on the logarithms of the jnd(I) to make sure that the variance was approximately independent of the size of the jnd(I) (see Florentine, 1983; Florentine *et al.*, 1987). Therefore, the average jnd(I) was calculated as the geometric mean of the individual data in decibels. An analysis of variance (ANOVA) for repeated measures was used to examine the statistical significance of the effects. Differences were considered significant when the tests indicated a probability less than 0.05.

II. EXPERIMENTS

A. Experiment I: Intensity discrimination as a function of shape factor and sensation level

The carrier frequency was 1 kHz. The level of the pink masking noise was set at 35 dB(A) SPL. The masked detection threshold of the stimuli was essentially constant as a function of shape factor (see the Appendix and Fig. 2.A3b for further discussion). The sensation levels of the stimuli were varied from 0 to 30 dB, in 10-dB steps.

Fig. 2.2 shows the discrimination threshold jnd(I) as a function of shape factor and sensation level for the individual subjects and the averaged discrimination thresholds across subjects. Error bars indicate the standard error of the mean.

The three listeners show similar behavior. At intermediate levels, i.e., at 10 dB SL for all subjects and at 20 dB SL for subjects JK and NS, jnd(I) varies as a function of the shape factor. When the shape factor is increased from 0.0375 to 0.15, intensity discrimination performance deteriorates (higher thresholds): at 10 dB SL, jnd(I) increases by a factor of 1.7 when the shape factor is quadrupled. When the shape factor is changed from 0.3 to 1.2, intensity discrimination performance improves (lower thresholds): at 10 dB SL, jnd(I) decreases by a factor of 1.4 when the shape factor is quadrupled. The maximum jnd(I) (poorest performance) occurs at 10 dB SL for shape factors of 0.15 and

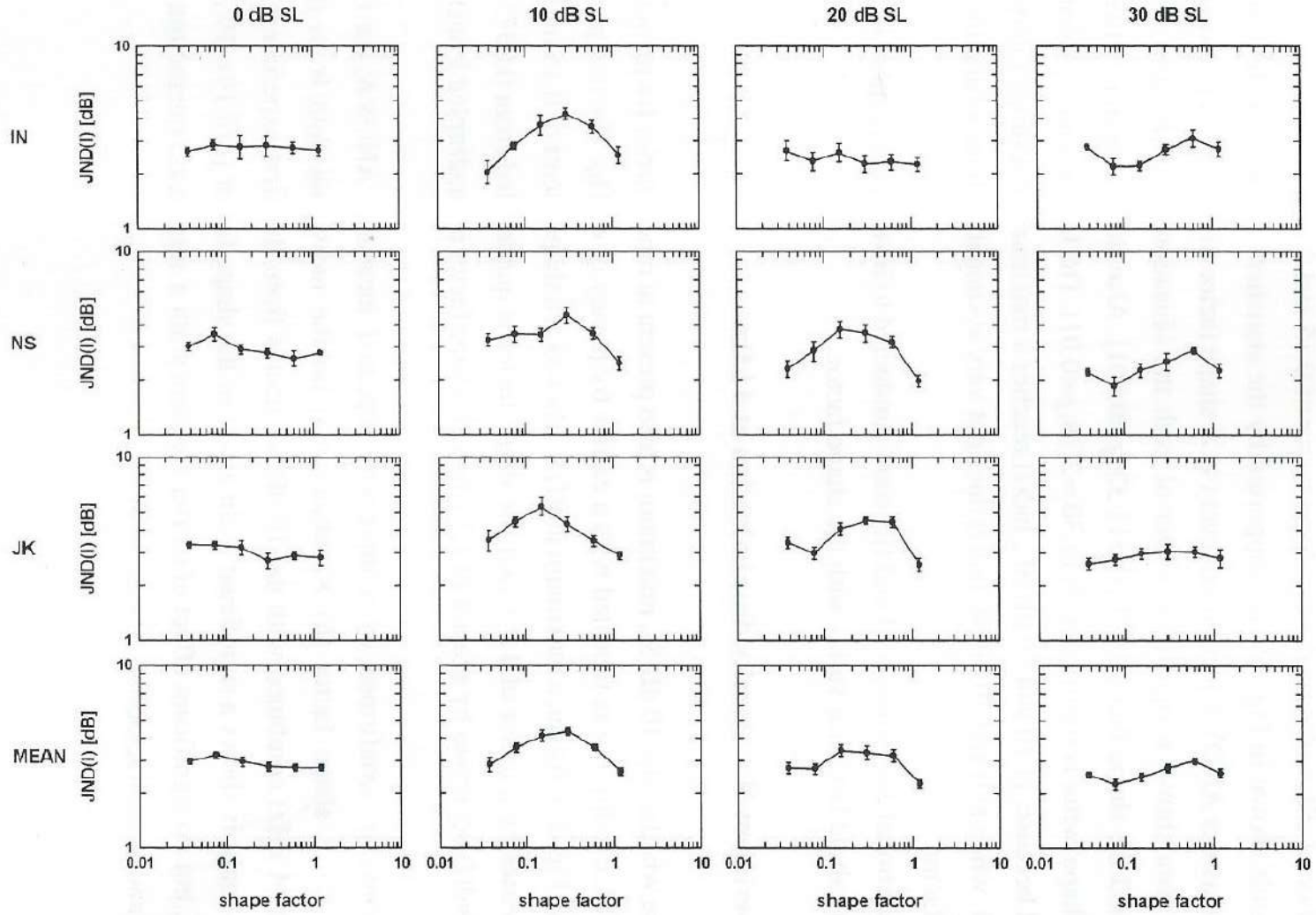


FIG. 2.2. jnd(I) for Gaussian-windowed 1-kHz tones plotted as a function of the shape factor, for the individual subjects (Rows 1–3) and averaged across subjects (Row 4). From the left to the right the sensation level is varied: 0, 10, 20, or 30 dB SL. Each point is the mean of six threshold estimates for one observer. The error bars show the standard error of the mean.

0.3. At lower and higher levels (0 and 30 dB, respectively) the jnd(I) does not vary with the shape factor. When the sensation level is increased from 0 to 10 dB, an increase in jnd(I) of about 1 dB is observed for a shape factor of 0.15 and 0.3.

The trends shown in Fig. 2.2 are supported by the statistical analysis. A three-way repeated measures ANOVA [sensation level (4) × shape factor (6) × subject (3)] on the individual data shows a significant effect of both the sensation level [$F(3,6)= 5.58$; $p=0.036$] and the shape factor [$F(5,10)=11.33$; $p<0.001$]. Also the interaction between level and shape factor is significant [$F(15,30)=2.78$; $p<0.01$]. The latter result is probably introduced because, at 10 and 20 dB SL, jnd(I) reaches a maximum at a shape factor of 0.15 or 0.3, while at 0 and 30 dB SL jnd(I) does not vary systematically as a function of the shape factor.

Two additional experiments (II and III) were conducted to investigate in more detail how the threshold behavior varies with the shape factor.

B. Experiment II: Intensity discrimination at 4 kHz

To examine whether the 10 dB SL maximum is also present at other carrier frequencies, the 10 dB SL condition was repeated with a carrier frequency of 4 kHz. The results are displayed in Fig. 2.3. Again, a maximum jnd(I) is reached at a shape factor of 0.15 or 0.3. jnd(I) increases by a factor of 1.7 when the shape factor is quadrupled from 0.0375 to 0.15, and jnd(I) decreases by a factor of 1.7 when the shape factor is quadrupled from 0.3 to 1.2.

The trends are confirmed by a three-way repeated measures ANOVA [carrier frequency (2) × shape factor (6) × subject (3)] on the individual data from this experiment (4 kHz) combined with the 10 dB SL results from the first experiment (1 kHz). The analysis shows a significant main effect of the shape factor [$F(5,10)=19.29$; $p<0.0001$], but no significant effect of carrier frequency nor a significant interaction of frequency and shape factor.

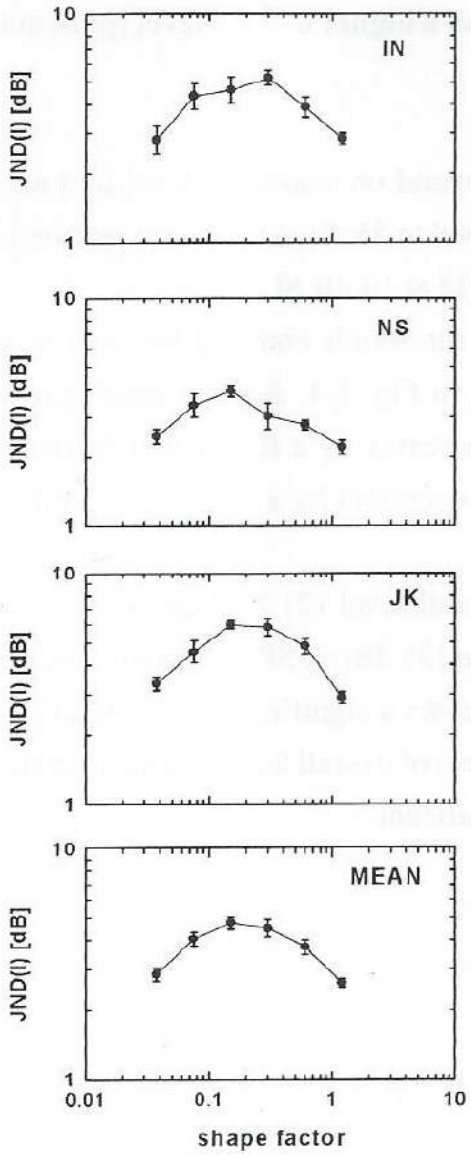


FIG. 2.3. jnd(I) for Gaussian-windowed 4-kHz tones plotted as a function of the shape factor. Sensation level: 10 dB; noise level: 35 dB(A) SPL. The upper three panels show the jnd(I) for the listeners separately. The lowest panel shows the mean result of the three listeners. Other details are the same as in Fig. 2.2.

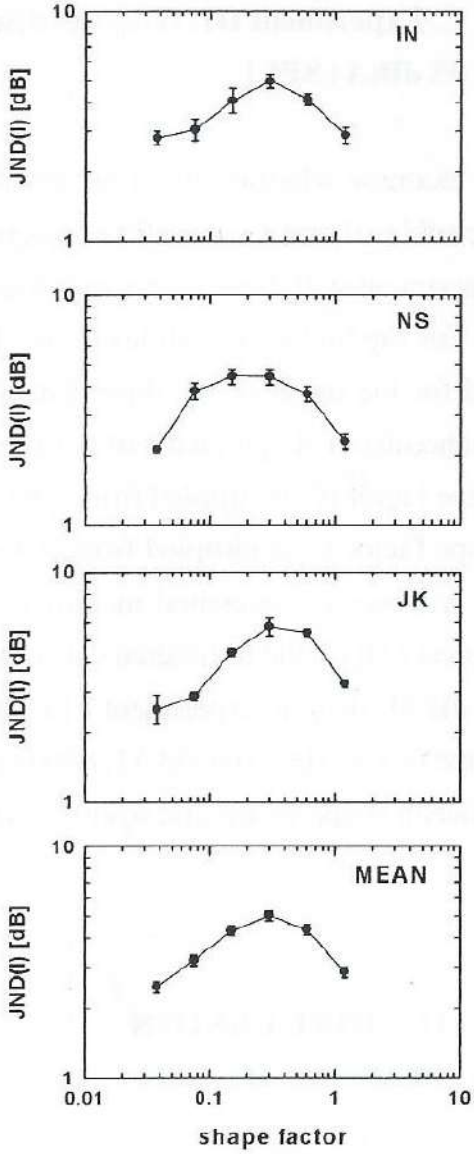


FIG. 2.4. jnd(I) for Gaussian-windowed 1-kHz tones plotted as a function of the shape factor. Sensation level: 10 dB; noise level: 55 dB(A) SPL. The upper three panels show the jnd(I) for the listeners separately. The lowest panel shows the mean result of the three listeners. Other details are the same as in Fig. 2.2.

C. Experiment III: Intensity discrimination for a higher overall level [pink noise: 55 dB(A) SPL]

To examine whether the observed trends really depend on sensation level and not on overall level, we increased the background noise level to 55 dB(A) SPL and repeated the experiment with Gaussian-windowed tones of 1 kHz at 10 dB SL.

For the higher overall level, the discrimination thresholds obtained for each subject and for the mean of the three listeners are shown in Fig. 2.4. Again a maximum was reached for a shape factor of 0.15 or 0.3. jnd(I) increases by a factor of 1.8 when the shape factor is quadrupled from 0.0375 to 0.15 and decreases by a factor of 1.7 when the shape factor is quadrupled from 0.3 to 1.2.

A three-way repeated measures ANOVA [overall level (2) × shape factor (6) × subject (3)] on the individual data of this experiment [55 dB(A) SPL] combined with the 10 dB SL data of experiment I [35 dB(A) SPL] shows a significant main effect of the shape factor ([$F(5,10)=33.51$; $p<0.000\,01$]. The effect of overall level and the interaction between shape factor and overall level are not significant.

III. DISCUSSION

The results show that the just-noticeable difference in intensity jnd(I) of Gaussian-windowed tones may vary as a function of the shape factor. For 1-kHz tones at sensation levels of 10 and 20 dB SL in 35 dB(A) SPL pink noise, jnd(I) reaches a maximum at a critical shape factor of 0.15 or 0.3 (see Fig. 2.2). At both lower and higher sensation levels, jnd(I) is relatively constant for different shape factors. For a 4-kHz carrier frequency, a similar variation in jnd(I) with the signal shape is obtained: again, at a shape factor of 0.15 or 0.3 a maximum is observed (see Fig. 2.3). Also, after increasing the overall level [noise level: 55 dB(A) SPL] the variation in jnd(I) persists (see Fig. 2.4).

In this study the spectro-temporal shape of the stimuli ranged from a relatively long-duration tone to a very short-duration click. As the signals vary from tone to click, two

processes occur: temporal shortening and spectral widening. First, as the signal decreases in duration and increases in bandwidth, the primary effect is temporal shortening. This causes an increase in the jnd(I) until the bandwidth reaches the critical band. At this point the second process, the increase in bandwidth, becomes important, serving to reduce the jnd(I). These two processes are addressed more or less separately in literature.

With respect to the effect of temporal shortening, Florentine (1986) and Buus and Florentine (1992) have done extensive research measuring intensity discrimination for pure tones as a function of duration. They found that intensity discrimination improves with increasing duration. Our results show the same behavior: at low levels (10 and 20 dB SL) jnd(I) decreases toward smaller shape factors, corresponding to longer durations. Also, quantitatively, the rate of improvement measured in this study agrees with the rate found by Florentine (1986) and Buus and Florentine (1992).

At some point, separating the effect of temporal shortening and the effect of the increase in bandwidth is not possible. Studying jnd(I) as a function of duration, Florentine (1986) omitted durations of 4 ms and less from the fitting procedure because these data deviated from a linear function [in a double logarithmic plot of jnd(I) versus duration]. She noted that this may have been due to the spectral splatter. Our data also show this effect, a flattening of the curve for small durations, at a shape factor between 0.15 and 0.3. Our explanation is, analogous to Florentine's remark, that at this point the bandwidth of the Gaussian-windowed signal exceeds the width of the auditory filter. From this point on, the *spectral* width of the signal determines the discrimination threshold, i.e., the process of increasing bandwidth becomes important.

With respect to the effect of the increase in bandwidth, Buus (1990) measured intensity discrimination as a function of bandwidth. He found that jnd(I) is independent of bandwidth when the stimulus bandwidth is less than the width of the auditory filter. For larger bandwidths, at low levels, a decrease in jnd(I) with increasing bandwidth was found. Our data also show this trend: jnd(I) decreases for shape factors larger than 0.3, at sensation levels of 10 dB and 20 dB.

In quantitative terms, an optimum detector predicts a decrease of jnd(I) by a factor of 2 (in decibels) when the bandwidth or the duration is quadrupled. In the multiple-window model the two processes of the variation in bandwidth and the variation in duration are combined in the variation of the number of f-t windows covered by the signal. Then, in the multiple-window model jnd(I) is expected to decrease by a factor of 2 when the

number of f-t windows is quadrupled. We found a decrease by a factor of 1.7 instead of 2, somewhat less than predicted by the multiple-window model. Florentine (1986) also found smaller improvements (a factor 1.5 when the duration was quadrupled). Possible explanations for this small deviation from the model improvement predictions are a reduced discriminability in the individual f-t windows as the number of f-t windows increases, a suboptimal combination of the information of the different windows (see also Buus and Florentine, 1992) or that the windows are not totally statistically independent.

The most important finding of this study is that, for intermediate sensation levels, the data qualitatively agree with the generalized multiple-window hypothesis put forward in the introduction. As a result, we can identify a "critical" shape factor, for which intensity discrimination performance is worst. This "critical" shape factor has a value between 0.15 and 0.3, both at a carrier frequency of 1 kHz and 4 kHz, at an overall level of 35 dB(A) SPL and 55 dB(A) SPL. So, in the proposed auditory spectro-temporal representation, a Gaussian-windowed sinusoid with a bandwidth of about 1/3 octave and an effective duration of about 4 ms at 1 kHz and 1 ms at 4 kHz (including effectively about four sinusoidal periods) can be considered an approximation of the "elementary" f-t window of the perceptually relevant auditory spectrogram. These values are in line with the idea of the critical band of the multiband excitation-pattern model (Florentine, 1986) and with the temporal constants of the multiple-look model, i.e., about 3 ms for 1-kHz tones and decreasing toward higher center frequencies (Viemeister and Wakefield, 1991).

Having discussed the main issue of the paper, i.e., the relation between jnd(I) and the shape factor, as observed at 10 or 20 dB SL, in terms of the multiple-window model, a few aspects of the data deserve some further discussion. (1) The masked detection threshold is virtually constant as a function of shape factor. (2) At 0 dB SL, jnd(I) does not depend on shape factor. (3) For the critical shape factor, jnd(I) increases about 1 dB when the sensation level increases from 0 to 10 dB SL. (4) At higher sensation levels (30 dB SL), intensity discrimination again is a constant as a function of the shape factor. First, the role of the internal noise versus the external noise in the multiple-window model will be addressed. This will help to clarify points 1 and 2. Then, points 3 and 4 will be discussed.

The noisy representation of intensities in the auditory system that underlies the observed intensity discrimination thresholds is formed as the sum of external and internal variance. The external variance is mediated by the external background noise added to

the signal in the experimental procedure. The internal variance is introduced in the auditory system itself, for example resulting from the variance in the neural coding process. If the signal-to-external-noise ratio is not too low, the internal noise dominates intensity discrimination performance. Following Weber's Law it is assumed that the variance due to internal noise is proportional to the signal energy. Thus the signal-to-internal-noise ratio is independent of the excitation level. Therefore, when the energy of the signal is distributed over several windows rather than concentrated within a single window, the "quality" within each individual window in terms of signal-to-internal noise ratio does not change. As a result, the combination of several windows will yield a better performance, for the internal noise is independent between windows. This forms the basis for the improvement predicted by the multiple-window model. These predictions are consistent with the results found at 10 and 20 dB SL.

However, at very low signal-to-external noise ratios the external noise dominates. Thus to clarify point 1 (masked detection thresholds) and point 2 [jnd(I) at 0 dB SL], the role of the external noise needs to be discussed. Contrary to the internal noise, the external noise in each window is signal *in*dependent. Therefore, when the energy of the signal is distributed over several f-t windows, the signal-to-external-noise ratio in each window decreases. This poorer quality in each individual window is counterbalanced by the combination of the information across several windows. The net effect is a constant threshold as a function of the number of f-t windows for the optimum detector. Thus the explanation for the constant masked thresholds (point 1) and the constant jnd(I) at 0 dB SL (point 2) is a trade off between the increase in the number of f-t windows covered by the signal and the decrease in the signal-to-external-noise ratio in each individual window. This might imply that masked detection thresholds cannot be used to assess the shape of the auditory f-t window.

The third point to be addressed is the observed increase in jnd(I) when the sensation level is increased from 0 to 10 dB. We believe that this is due to a two-stage strategy listeners will use in an intensity discrimination task at 0 dB SL: a detection stage followed by a discrimination stage. At 0 dB SL not all stimuli are detectable; the stimulus with the incremented amplitude has a higher probability of being detected. Detecting a stimulus in a particular interval is a one-interval process; discriminating among the stimuli is a three-interval process. Thus due to the difference in memory load (Durlach and Braida, 1969), assuming that listeners benefit from the detection cue at 0 dB SL seems reasonable

(see also the Appendix). As a result jnd(I) is smaller at 0 dB SL than at 10 dB SL. Because most studies regarding jnd(I) as a function of level report a decreasing jnd(I) as a function of level, our results, the increase in jnd(I) when the sensation level increases from 0 to 10 dB, might seem a little unexpected. However, in most studies (see, e.g., Jesteadt *et al.*, 1977, Florentine, 1983; Florentine *et al.*, 1987, Ozimek and Zwislocki, 1996) the lowest sensation level at which the jnd(I) is measured is 5 or 10 dB SL; At this level the effect of the detection strategy has probably disappeared. The only study known by the authors that measured the jnd(I) at 0 dB SL was a study by Hanna *et al.* (1986). Unfortunately, their results can neither confirm nor disprove our results.

Regarding the last point (i.e., 4), spread of excitation is important. In the spectral domain, it is well known that spread of excitation, i.e., the growth of the excitation pattern with increasing level, occurs. As a result, jnd(I) is independent of bandwidth at high sensation levels. This effect was found by, for instance, Buus (1990) and is also accounted for by the multiband excitation-pattern model (Florentine and Buus, 1981; Buus and Florentine, 1994). In the multiple-window approach, spread of excitation is anticipated both in the temporal and in the spectral domain: The higher the sensation level, the larger the area on the f-t plane excited by the signal. Therefore, at 30 dB SL, probably even for the critical shape factor the internal signal representation may already cover many elementary f-t windows. This may explain why jnd(I) becomes independent of the shape factor at higher levels.

To substantiate the qualitative arguments of the multiple-window idea, and the role of external and internal noise as described in the preceding paragraphs, a simple detection and discrimination model was developed. We refer to the Appendix for a description of the model. The aim of the model is to simulate the trends observed in the data: the dependence of the discrimination threshold on the shape factor at 10 dB SL, whereas at 0 dB SL the discrimination threshold is a constant as a function of shape factor; the constant detection threshold as a function of shape factor; the slight increase in jnd(I) at a shape factor of 0.3 when the sensation level increases from 0 to 10 dB SL. The simulated trends (see the Appendix) agree with the observed trends in the data.

The results of this study point to an auditory f-t window with a constant relative bandwidth and a duration inversely related to frequency: at low frequencies the spectral width of the f-t windows is small and the duration long, whereas at high frequencies the spectral width of the f-t windows is broad and the duration short. This perceptually

4 The effect of intensity perturbations on speech intelligibility for normal-hearing and hearing-impaired listeners

Hearing-impaired listeners are known to suffer from reduced speech intelligibility in noise, even if sounds are above their hearing thresholds. This study examined the possible contribution of reduced acuity of intensity coding to this problem. The "distortion-sensitivity model" was used: the effect of reduced acuity of auditory intensity coding on intelligibility was mimicked by an artificial distortion of the speech intensity coding, and the sensitivity to this distortion for hearing-impaired listeners was compared with that for normal-hearing listeners. Stimuli (speech plus noise) were wavelet coded using a Gaussian wavelet (1/4 octave bandwidth). The intensity coding was distorted by multiplying the modulus of each wavelet coefficient by a random factor. Speech-reception thresholds (SRTs) were measured for various degrees of intensity perturbation. Hearing-impaired listeners were classified as suffering from suprathreshold deficits if intelligibility of undistorted speech was worse than predicted from audibility by the Speech Intelligibility Index model (ANSI, 1997). Hearing-impaired listeners without suprathreshold deficits were as sensitive to the intensity distortion as the normal-hearing listeners. Hearing-impaired listeners with suprathreshold deficits appeared to be less sensitive. Results indicate that reduced acuity of auditory intensity coding may be a factor underlying reduced speech intelligibility for the hearing impaired.

Submitted to the Journal of the Acoustical Society of America

at 10 dB SL will be simulated. In the simulation this translates into the dependence of the discrimination threshold on the number of elementary f-t windows covered by the signal. Then, analogous to the points addressed in Sec. 2.III, the following trends will be simulated: (1) the detection threshold as a function of the number of f-t windows covered by the signal; (2) the discrimination threshold as a function of the number of windows at 0 dB SL; (3) for one f-t window, the discrimination threshold at 0 and 10 dB SL.

In the simulations the 3AFC two-down one-up adaptive procedure (see Sec. 2.IC) is adopted: in each trial three intervals are presented; a decision algorithm decides which interval contains the signal in case of detection or the incremented signal in case of discrimination. Thus in the model a human observer is mimicked and the simulated thresholds can be compared directly to the experimental data.

All signals have total energy E. If a signal covers just one f-t window of the auditory system, this f-t window contains the total energy E. If a signal extends over a number of N f-t windows, the N f-t windows contain each $1/N$ part of E. In Table 2.I a rough estimate of the number of f-t windows corresponding to the stimuli used in the experiments can be found. This estimate is based on an f-t window with a shape factor of 0.23 (about the "critical" shape factor), corresponding to a Gaussian-windowed stimulus with a bandwidth of 1/3 octave. The external noise that enters each f-t window is modeled as Gaussian noise with spectral density N_0. This noise having a random phase and an amplitude taken from a Rayleigh distribution is added to the signal. In the model, the external noise of the different f-t windows is assumed to be uncorrelated.

A. Detection

In the simulated detection experiments, one interval contains the signal plus external noise and the other two contain only external noise. In Fig. 2.A1a a scheme of the detection model is plotted. If the signal covers more than one f-t window (in case of small and large shape factors), the energy of the signal within an interval is divided over the proper number of f-t windows. In a detection task where stimuli are not always audible, assuming that the auditory system is unable to focus exactly on the f-t windows covered by the signal seems reasonable. Therefore, 50 f-t windows are considered for all shape factors, comparable to, for example, an integration time of 200 ms (50 times 4 ms). On this internal auditory representation, detection decisions are based. Detection performance

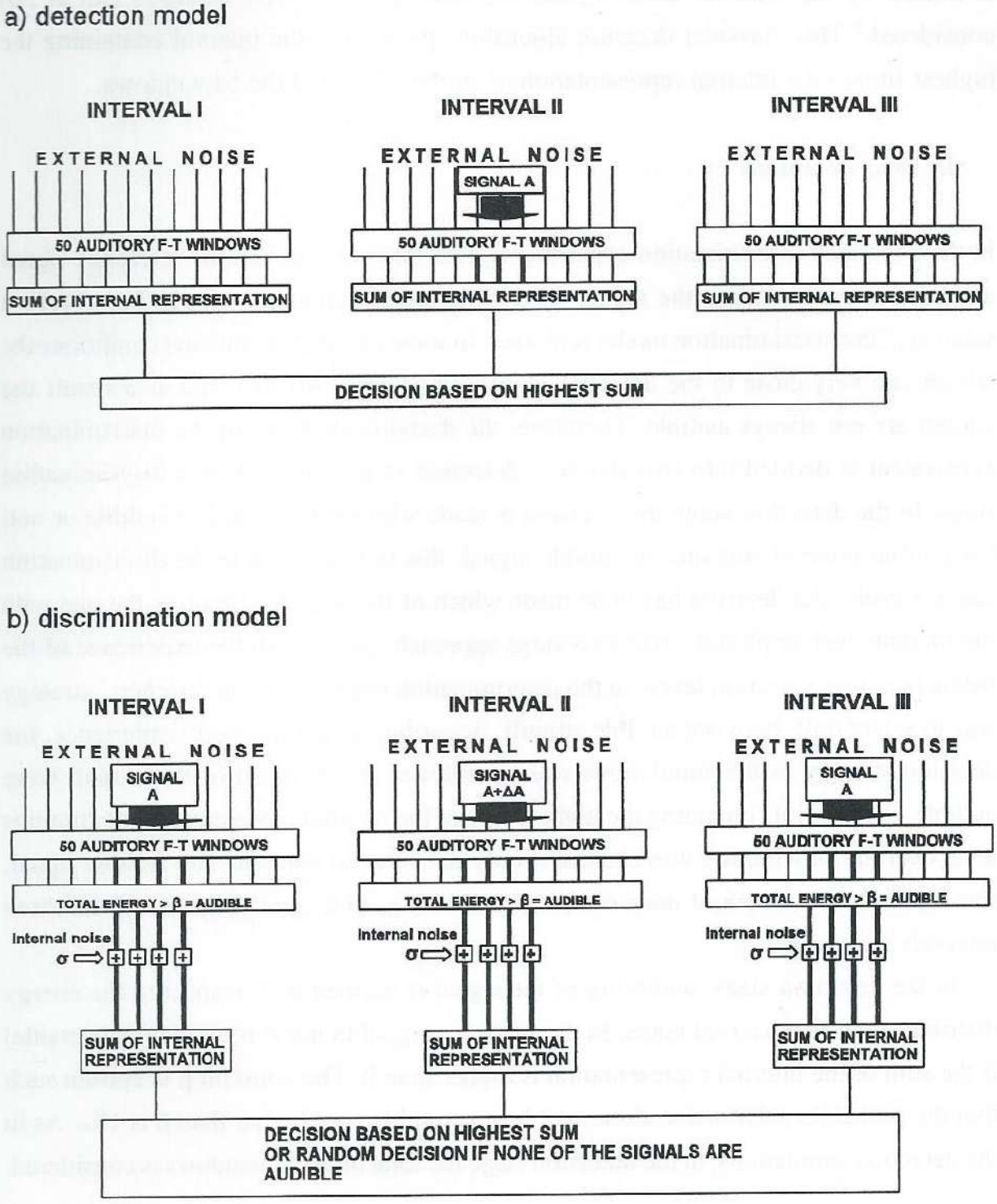


FIG. 2.A1. A schematic representation of the detection (a) and discrimination (b) model.

is limited by the external noise N_0 and the total number of f-t windows (set at 50) considered.[4] This classical decision algorithm "picks out" the interval containing the highest sum of the internal representation of stimulus level of the f-t windows.

B. Discrimination

In the simulated discrimination experiments, two intervals contain the reference signal and one interval contains the signal with the incremented amplitude. In Fig. 2.A1b a scheme of the discrimination model is plotted. In some of our experimental conditions the stimuli are very close to the detection threshold (0 or 10 dB SL), and as a result the stimuli are not always audible. Therefore, the decision strategy for the discrimination experiment is divided into two stages: a detection stage followed by a discrimination stage. In the detection stage the decision is made whether the signal is audible or not. Only, if an interval contains an audible signal, this is forwarded to the discrimination stage. Finally, the decision has to be made which of the audible stimuli is the one with the incremented amplitude. This two-stage approach agreed with the experience of the listeners at low sensation levels in the discrimination experiment: the listeners' strategy was to select only between audible stimuli. According to the listeners' experience, the decision strategy in the simulations was as follows: If two or all of the stimuli were audible, the interval containing the highest sum of the internal representation of stimulus level over the f-t windows was chosen; if only one interval contained an audible signal, this interval was chosen; if none of the stimuli was audible, randomly one of the three intervals was picked.

In the detection stage, audibility of the signal is defined with respect to the energy distribution of the external noise. In the model, a signal in noise is audible (detectable) if the sum of the internal representation is higher than β. The constant β is chosen such that the probability that noise alone will have a total energy higher than β is 1‰. As in the detection simulations, in the detection stage the total of 50 f-t windows is considered.

---

[4] The fact that the total energy of 50 windows was considered for all stimuli is an essential part of the detection model, and will affect the detection threshold as a function of the number of f-t windows.

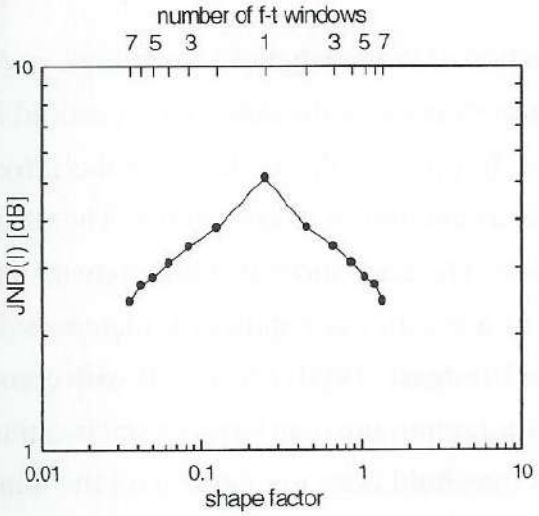


FIG 2.A2. jnd(I) as a function of the shape factor at 10 dB SL. An estimate of the number of f-t windows corresponding to the shape factors is shown on the top axis.

In the discrimination stage where the signal is always audible, assuming that the listener can focus exactly on the f-t windows covered by the signal seems reasonable. Therefore, only the f-t windows containing the signal are considered. "Coding" noise is added to the internal representation. From the literature (see, e.g., Buus and Florentine, 1991) it is known that, in discrimination tasks, the sensitivity d' is roughly proportional to the difference limen in intensity: $d'=k*\log_{10}((E+\Delta E)/E)$. Therefore, the variance of the internal "coding" noise component in the discrimination stage was taken to be proportional to the energy of the signal (constant variance in dB, Weber's Law): The noise was taken from a Gaussian distribution with a standard deviation σ. Considering the range of the jnd(I) of our results, $\sigma = 4$ dB was taken as a reasonable value. The internal noise is uncorrelated across the f-t windows [see Durlach *et al.* (1986)].

C. Results of the simulations

The simulated discrimination threshold as a function of the shape factor at 10 dB SL is plotted in Fig. 2.A2. On the top axis of Fig. 2.A2 the number of f-t windows used to simulate the different shape factors is shown. The shape factors and the corresponding estimate of the number of f-t windows can also be found in Table 2.I. At 10 dB SL, jnd(I) has a maximum for the critical shape factor or, alternatively, for one f-t window. jnd(I) decreases for smaller and larger shape factors, or, alternatively, as the number of f-t

windows increases. These trends are also observed in the data (see Figs. 2.2, 2.3, and 2.4).

The simulated detection threshold as a function of the number of f-t windows covered by the signal is plotted in Fig. 2.A3a. The figure shows that the detection threshold E/N_0 is independent of the number of f-t windows. In Fig. 2.A3b, the mean of the informal detection threshold data at 1 kHz is plotted as a function of shape factor. The data are expressed in decibels *re:* an arbitrary reference. The data show a slight increase in the detection threshold as the shape factor and, as a result, the bandwidth increases. This trend was also observed by Van den Brink and Houtgast (1990) for signals with constant spectro-temporal area. Because no maximum (nor minimum) can be observed in our data, it is concluded that, essentially, the detection threshold does not depend on the number of f-t windows covered by the signal. Experimentally observed and simulated trends agree.

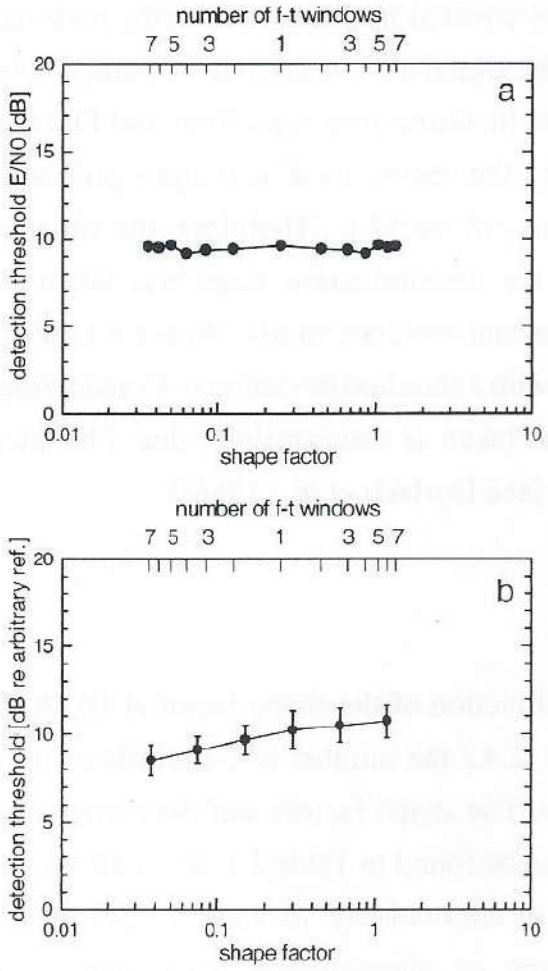


FIG. 2.A3. (a) The model-predicted detection threshold as a function of the shape factor. An estimate of the number of f-t windows corresponding to the shape factors is shown on the top axis. (b) The mean of the informal detection threshold data at 1 kHz as a function of shape factor. The data are expressed in decibels re an arbitrary reference.

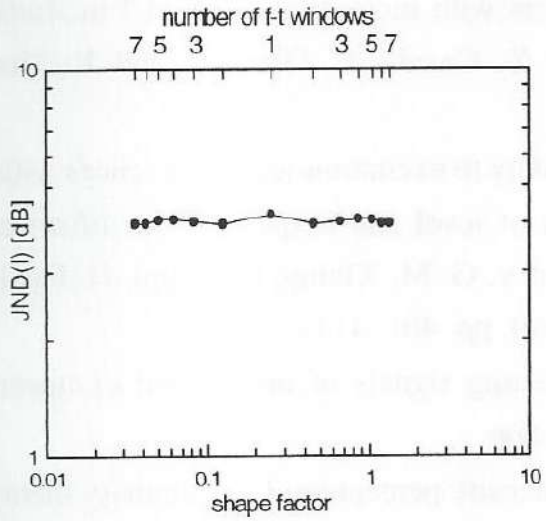


FIG. 2.A4. jnd(I) as a function of shape factor at 0 dB SL. The corresponding number of f-t windows is shown on the top axis.

The simulated discrimination threshold as a function of the shape factor at 0 dB SL is shown in Fig. 2.A4. On the top axis of the figure the number of f-t windows used to simulate the different shape factors is shown. Please see also Table 2.I. The jnd(I) at 0 dB SL does not depend on the shape factor, or, alternatively, the number of f-t windows. The simulated trends agree with the data (see Fig. 2.2). Comparing Fig. 2.A2 (10 dB SL) and Fig. 2.A4 (0 dB SL), it can be seen that for the "critical" shape factor (or one f-t window) the jnd(I) increases with about 1 dB when the sensation level increases from 0 to 10 dB SL. This trend was also observed in the data (see Fig. 2.2).

REFERENCES

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

Buus, S. (**1990**). "Level discrimination of frozen and random noise," J. Acoust. Soc. Am. **87**, 2643–2654.

Buus, S., and Florentine, M. (**1991**). "Psychometric functions for level discrimination," J. Acoust. Soc. Am. **90**, 1371–1380.

Buus, S., and Florentine, M. (**1992**). "Possible relation of auditory-nerve adaptation to slow improvement in level discrimination with increasing duration," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Démany, and K. Horner (Pergamon, New York), pp. 279–288.

Buus, S., and Florentine, M. (**1994**). "Sensitivity to excitation-level differences within a fixed number of channels as a function of level and frequency," in *Advances in Hearing Research*, edited by G. A. Manley, G. M. Klump, C. Köppl, H. Fastl, H. Oeckinghaus (World Scientific, Singapore), pp. 401–414.

Dai, H., and Wright, B. A. (**1995**). "Detecting signals of unexpected or uncertain durations," J. Acoust. Soc. Am. **98**, 798–806.

Durlach, N. I., and Braida, L. D. (**1969**). "Intensity perception. I. Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.

Durlach, N. I., Braida, L. D., and Ito, Y. (**1986**). "Towards a model for discrimination of broadband signals," J. Acoust. Soc. Am. **80**, 63–72.

Florentine, M. (**1983**). "Intensity discrimination as a function of level and frequency and its relation to high-frequency hearing," J. Acoust. Soc. Am. **74**, 1375–1379.

Florentine, M. (**1986**) "Level discrimination of tones as a function of duration," J. Acoust. Soc. Am. **79**, 792–798.

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

Florentine, M., Buus, S., and Mason, C. R. (**1987**). "Level discrimination of tones as a function of level and frequency from 0.25 to 16 kHz," J. Acoust. Soc. Am. **81**, 1528–1541.

Florentine, M., Fastl, H., and Buus, S. (**1988**). "Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking," J. Acoust. Soc. Am. **84**, 195–203.

Gabor, D. (**1947**). "Acoustical quanta and the theory of hearing," Nature (London) **159**, 591–594.

Gerken, G. M., Bhat, V. K. H., and Hutchison-Clutter, M. H. (**1990**). "Auditory temporal integration and the power-function model," J. Acoust. Soc. Am. **88**, 767–778.

Hanna, Th. E., von Gierke, S. M., and Green, D. M. (**1986**). "Detection and intensity discrimination of a sinusoid," J. Acoust. Soc. Am. **80**, 1335–1340.

Jesteadt, W., Wier, C. C., and Green, D. M. (**1977**). "Intensity discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 169–177.

Levitt, H. (**1971**). "Transformed up–down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 1519–1527.

Ozimek, E., and Zwislocki, J. J., (**1996**). "Relationships of intensity discrimination to sensation and loudness levels: Dependence on sound frequency," J. Acoust. Soc. Am. **100**, 3304–3320.

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

Rioul, O, and Vetterli, M. (**1991**). "Wavelets and signal processing," IEEE Signal Proc. Mag., October, 14–38.

Scharf, B., and Buus, S. (**1986**). "Audition I: Detection and discrimination," in *Handbook of Perception and Human Performance*, edited by K. Boff (Wiley, New York).

Stewart, G. W. (**1931**). "Problems suggested by an uncertainty principle in acoustics," J. Acoust. Soc. Am. **2**, 325–329.

Versfeld, N. J., Dai, H., and Green, D. M. (**1996**). "The optimum decision rules for the oddity task," Percept. Psychophys. **58**, 10–21.

Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.

3 Wavelet analysis

Since wavelet analysis plays an important role in this thesis, a short explanation about this joint time-frequency analysis will be given in this chapter. The similarities between wavelet analysis and auditory analysis will be addressed. This chapter will go into the basic parameters of the wavelet tool that will be used in the following chapters.

INTRODUCTION

In this thesis, wavelet coding is used as a tool for studying the auditory system. Different perspectives with respect to this topic are possible. Since this thesis deals with acoustic signals, a sound processing point of view will be taken. Then, wavelet analysis can be considered an analysis in which both temporal and spectral information of the signal are obtained, just as in auditory sound analysis. In this chapter, some general aspects of wavelet analysis will be explained. Also, the differences between wavelet analysis and short-time Fourier analysis will be discussed (Sec. 3.II). We will focus on the similarities between auditory analysis and wavelet analysis, and on how the parameters of the wavelet analysis can be tailored to the auditory system (Sec. 3.III). The result is a perceptually relevant sound coding, that will be called auditory wavelet coding (Sec. 3.V). This auditory wavelet coding will be used as a front-end signal processing tool to

study the auditory system in Chapters 4 and 5. When no references are given, the text of this chapter is based on Rioul and Vetterli (1991), Vetterli and Kovacevic (1995), and Strang and Nguyen (1996).

I. BASICS OF WAVELET ANALYSIS

A. Wavelets

As the name suggests, wavelet analysis is an expansion by means of wavelets. Wavelets are little waves. In Fig. 3.1a, an example of a wavelet is shown. Wavelets are oscillatory and decay to zero quickly. In acoustics, a wavelet is equivalent to a time-frequency window. It is localized in time and in frequency and does not have a DC component. All wavelets in a particular wavelet analysis are based on a fundamental prototype analysis function, i.e., the mother wavelet. Many different mother wavelets are possible, a constraint being that they should integrate to zero. Different wavelets within one analysis scheme are scales and shifts of this mother wavelet φ_M:

$$\varphi(t)=\frac{1}{\sqrt{a}}\varphi_M\left(\frac{t-t_0}{a}\right) \qquad a\in R^+,\ t_0\in R \tag{3.1}$$

in which a is the scaling parameter, t_0 is the shifting parameter, and $1/\sqrt{a}$ normalizes the energy of the wavelets.

In Fig. 3.1, examples of scales and shifts of the mother wavelet are shown. Scaling is compression or stretching in time of the mother wavelet (Fig. 3.1b). The smaller the scale, the more compressed the wavelet. By scaling, wavelets with different positions along the spectral axis are obtained. [In the example of Fig. 3.1, the carrier frequency f_0 of the wavelet is inversely proportional to the scale ($\propto 1/a$).] Scaling also affects the temporal and spectral width of the wavelet. This aspect will be discussed in the next section. For all wavelets, the number of oscillations within the temporal envelope is constant. By a shift of the mother wavelet, a wavelet with a different position along the temporal axis is obtained (Fig. 3.1c). Thus, by scales and shifts of the mother wavelet the whole time-frequency range can be spanned.

B. Scale

Several times, the term 'scale' was mentioned. Scale plays a central role in wavelet analysis. Signals are analyzed at different scales obtained by compression or stretching of the mother wavelet. By compression, a wavelet at a small scale is obtained. This contracted wavelet can be used to analyze detailed aspects, i.e., the high frequencies, in a signal. It is very localized in time, but less localized in frequency. Thus, at higher frequencies, the spectral resolution of wavelet analysis is reduced, but the temporal resolution is increased. By stretching the mother wavelet, a wavelet at a large scale is constructed. This wavelet has a long duration and can be used to analyze long-term trends, i.e., the low frequencies, in a signal. This wavelet is not so much localized in time, but very localized in frequency. Thus, at lower frequencies, spectral resolution is good, but temporal resolution is poor.

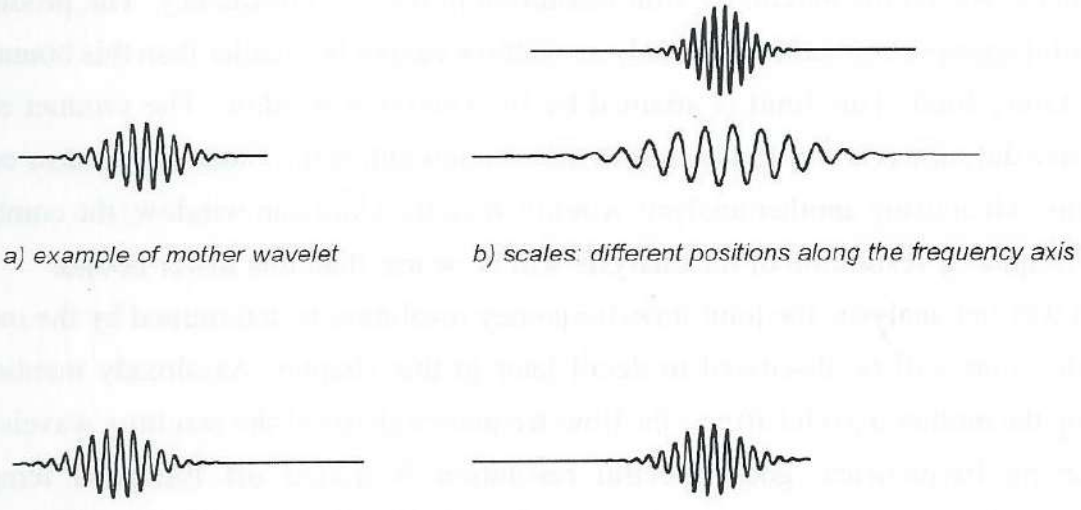


FIG. 3.1. Wavelets

In summary, wavelet analysis uses short windows at high frequencies and long windows at low frequencies. As a result, for increasing frequencies temporal resolution gets better, but spectral resolution gets poorer. The spectral resolution of a wavelet analysis is inversely proportional to the scale; it is proportional to the carrier frequency of the wavelets. Therefore, wavelet analysis has a constant *relative* spectral resolution. The zooming-in property, to analyze according to scale, is fundamental to wavelet analysis.

C. Uncertainty principle

Important characteristics of a time-frequency analysis are its temporal and spectral resolution. The uncertainty principle states that it is not possible to get an arbitrarily good resolution both in time and in frequency (see Landau and Polak, 1964). Maximum spectral resolution is obtained by long-term Fourier analysis, but then, no temporal information is available. Maximum temporal resolution is provided by the time signal itself, but then, no spectral information is available. By application of a time-frequency window both temporal and spectral information of the signal can be obtained. However, a bound exists on the maximum joint resolution in time and frequency. The product of temporal and spectral width of an analysis window cannot be smaller than this bound, the uncertainty limit. This limit is attained by the Gaussian window. The product of the effective duration (Gabor, 1947) and effective bandwidth of the Gaussian window equals 1. Thus, when using another analysis window than the Gaussian window, the combined time-frequency resolution of the analysis will be worse than this lower bound.

In wavelet analysis, the joint time-frequency resolution is determined by the mother wavelet, that will be discussed in detail later in this chapter. As already mentioned, scaling the mother wavelet affects the time-frequency shape of the resulting wavelet: for increasing frequencies, good spectral resolution is traded off for good temporal resolution. However, the product of bandwidth and duration is constant and always larger than or equal to the uncertainty limit.

D. Scalogram

The scalogram is the time-frequency representation of the energy of a signal analyzed by means of wavelets (Rioul and Flandrin, 1992). It is similar to the spectrogram of the short-time Fourier transform (to be discussed in the next section). A wavelet expansion results in wavelet coefficients. A wavelet coefficient has a modulus and a phase. It corresponds to a specific wavelet in the expansion. The square of the modulus of the coefficient represents the energy of the signal at a particular time and frequency. In Fig. 3.3b, an example of a scalogram is plotted. Presented is the energy distribution over time and frequency of a signal consisting of a sinusoid added to a spike. The scalogram is shaded proportionally to the energy. The time-frequency plane is covered by so-called time-frequency tiles. The term time-frequency tile of a particular wavelet is used to designate the time-frequency position and width of that analysis function (Herley *et al.*, 1993). In the scalogram, the tiles are symbolized by rectangles. In Fig. 3.2 a more realistic representation of an elementary tile corresponding to a Gaussian wavelet is shown. The tiles of the scalogram represent the time-frequency shape of the analyzing wavelets, expressing the temporal and spectral resolution. As can be observed in Fig. 3.3b, the spectral resolution is not constant nor is the temporal resolution. However, the area, i.e., the product of temporal and spectral width of the wavelets, is constant.

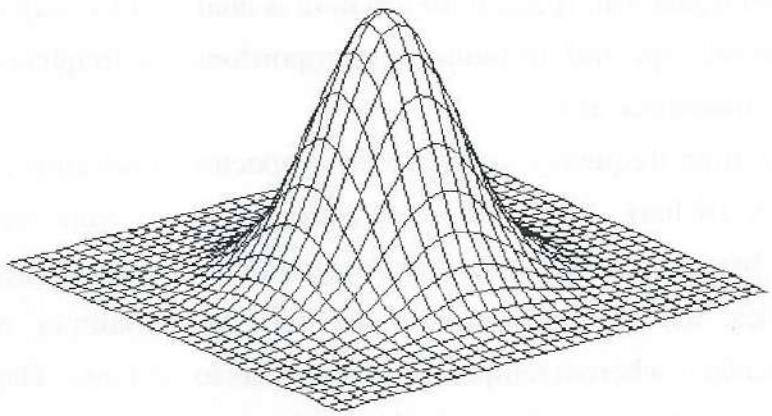


FIG. 3.2. An elementary tile corresponding to a Gaussian wavelet.

II. THE WAVELET TRANSFORM VS. THE SHORT-TIME FOURIER TRANSFORM

Classically, the short-time Fourier transform is used for time-frequency analyses. The wavelet transform has recently gained interest as an analysis tool for acoustic signals. The short-time Fourier transform is a Fourier transform on short-time segments of a signal. First, the signal is windowed using a fixed temporal window. Then, each sound segment is transformed to the Fourier domain using a basis of sines and cosines. The spread of the energy over time and frequency is represented in a spectrogram.

Just like the wavelet transform, the short-time Fourier transform is a joint time-frequency analysis, but the analysis windows of the short-time Fourier transform and the wavelet transform are very different. In Fig. 3.3 this is illustrated by a schematic representation of a scalogram and a spectrogram. In a wavelet analysis, different frequencies are obtained by scaling of the mother wavelet. By compression, the envelopes of the wavelets are narrowed for higher frequencies. The number of oscillations under a window is constant. The short-time Fourier transform uses analysis functions of constant duration. As a result, for higher frequencies, an increasing number of oscillations are present under the envelope. In a short-time Fourier transform, spectral information is analyzed by sines and cosines windowed by a fixed temporal window. As a result, the spectral resolution of the short-time Fourier transform is independent of time and frequency. For the wavelet transform, spectral information is analyzed by scales of the mother wavelets. As a result, spectral resolution is proportional to frequency; it is constant on a logarithmic frequency axis.

Auditory coding is a time-frequency coding with a spectral resolution roughly proportional to frequency (Scharf, 1970). Wavelet coding has this same property. Moreover, the results of Chapter 2 show that the initial stages of the auditory system have wavelet-like characteristics: the spectral width of the peripheral auditory window increases for higher frequencies, whereas temporal width appears to decrease. Therefore, wavelet analysis provides an interesting alternative to the short-time Fourier Transform to model and understand time-frequency coding of the auditory periphery.

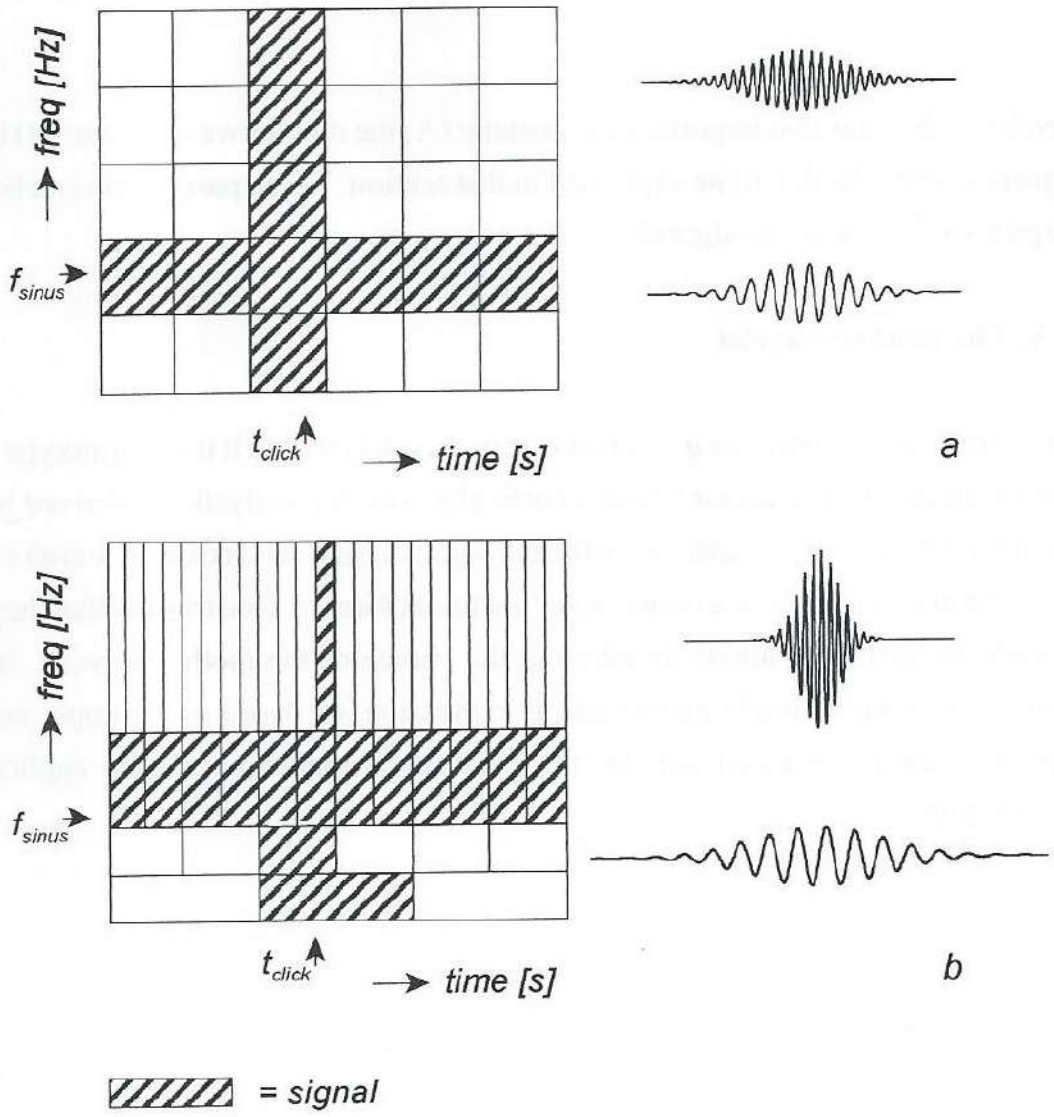


FIG. 3.3. A schematic illustration of the energy distribution of a click at time t_{click} added to a sinusoid of frequency f_{sinus}. In panel (a) the result of a short-time Fourier analysis is plotted (spectrogram), and in panel (b) the result of a wavelet analysis (scalogram) is plotted. On the right side of the figures, examples of the corresponding analysis windows of the short-time Fourier analysis and wavelet analysis are shown.

III. IMPORTANT PARAMETERS OF WAVELET CODING

Wavelet coding has two important parameters: (A) the mother wavelet, and (B) the time-frequency tiling. Both will be explained in this section. These parameters can be used to design a wavelet analysis algorithm.

A. The mother wavelet

The mother wavelet plays an important role in wavelet coding. It is the prototype analysis function from which all wavelets in a particular wavelet analysis are derived by scales and shifts. Unlike Fourier analysis, which is based on sines and cosines, wavelet analysis can use mother wavelets of a rather wide functional form. A constraint is that they should integrate to zero. This allows freedom in the choice of the mother wavelet. It can be smooth, based on a simple mathematical expression, or based on a simple associated filter. The mother wavelet can be made to fit or model a specific application or phenomenon.

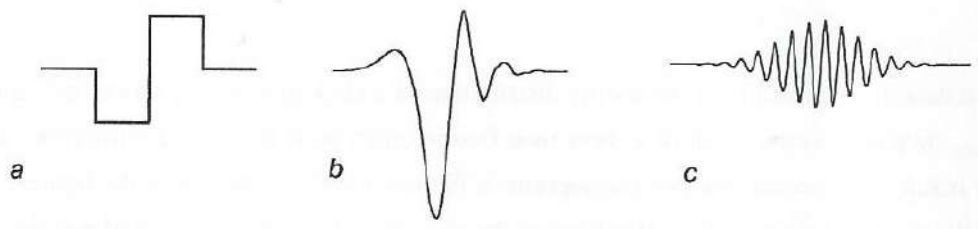


FIG. 3.4. Mother wavelets: (a) Haar wavelet; (b) Daubechies 7 wavelet; (c) Gaussian wavelet.

An important class of mother wavelets is that with compact support in time. Compact support means that the mother wavelet is zero outside a certain interval. Numerically, this is an attractive quality. The Daubechies wavelets, named after their inventor, have compact support in time. The first and most simple wavelet out of the class of Daubechies wavelets, i.e., Daubechies 1, was already known longer. It is also called the Haar wavelet (see Fig. 3.4a). The Haar wavelet was named after Haar who, in 1910, was the first one to construct a basis, not by sines and cosines, but by scales and shifts of this step function. Having extremely compact support the Haar wavelet is very localized in time, but it is not well localized in frequency. Its spectrum has many large sidelobes. In Fig. 3.4b, the Daubechies 7 wavelet is shown. Daubechies wavelets with a higher number are less localized in time, but better localized in frequency. The mother wavelet used in this thesis, i.e., the Gaussian wavelet, does not have compact support in time, but it is spectrally very smooth (Fig. 3.4c).

Different mother wavelets have different time-frequency shapes. This shape determines the spectral and temporal resolution of the wavelet analysis. Therefore, the choice of the mother wavelet has great impact on the display of the time-frequency characteristics of the signal under investigation. As already mentioned, the uncertainty principle puts a lower bound on the product of temporal and spectral resolution. However, above this bound one is free to choose an adequate joint time-frequency resolution.

As explained before, the periferal auditory time-frequency window looks a lot like that of a wavelet. The question remains what mother wavelet most closely resembles the auditory time-frequency window. The results of Chapter 2 suggest that a Gaussian-windowed sinusoid with a shape factor between 0.15 and 0.3 roughly matches the auditory time-frequency window. Thus, a Gaussian mother wavelet may be a reasonable choice for a wavelet coding that is similar to peripheral auditory coding. The Gaussian wavelet is a complex sinusoidal carrier with a Gaussian envelope. It is described by

$$s(t) = \sqrt{\alpha f_0}\, \exp(i2\pi f_0(t-t_0))\, \exp(-\pi(\alpha f_0(t-t_0))^2) \quad , \tag{3.2}$$

in which f_0 is the carrier frequency, α is the shape factor, and $\sqrt{\alpha f_0}$ normalizes the energy of the analysis function. As shown in equation 3.1, wavelets are constructed by scales and shifts of this mother wavelet. Since f_0 is inversely proportional to scale, different scales are obtained by varying f_0. Different shifts are obtained by varying t_0. This

frequency-time window has an effective bandwidth of $\Delta_f=\alpha f_0$ and an effective duration of $\Delta_t=1/(\alpha f_0)$ (Gabor, 1947). A numerical drawback of the Gaussian mother wavelet is that it does not have compact support[1] in time. However, its fundamental advantage is smoothness both in time and in frequency. In contrast to many other mother wavelets, it does not have spectral sidelobes. Its bandwidth and duration can be adjusted to fit the bandwidth and duration of the auditory time-frequency window.

Another option would be to use an asymmetric time-frequency window, because the auditory time-frequency window is probably not symmetrical (see, e.g., Irino and Patterson, 1996). For example, the asymmetric gammatone may provide a better approximation of the auditory time-frequency window than the symmetric Gaussian wavelet (Irino and Patterson, 1997). A drawback of the gammatone is that it is numerically less efficient than the Gaussian wavelet, because one side of the gammatone envelope decays more slowly than the other one. An appealing property of the Gaussian wavelet is its similarity in time and frequency. Moreover, it satisfies minimal uncertainty in the joint time-frequency representation. Since the Gaussian wavelet can be considered a first order approximation of the gammatone and to keep computations simple, in this thesis, a Gaussian mother wavelet was chosen for auditory wavelet coding. In Chapters 4 and 5, the effective bandwidth of the Gaussian wavelet is set to ¼ octave [roughly equal to the auditory critical band (Scharf, 1970)]. This corresponds with a shape factor α of 0.1735. As a result, the effective duration of the frequency-time window is 5.76 ms at 1 kHz (1.44 ms at 4 kHz). The effective number of sinusoidal periods within the Gaussian envelope equals 5.8 (i.e., $1/\alpha$).

B. Time-frequency tiling

The scalogram of Fig. 3.3b showed that the time-frequency plane is covered by tiles, i.e., time-frequency windows. The time-frequency tiling is related to the sampling in time and frequency of a wavelet expansion, because it indicates where wavelets are localized in time and frequency (Herley *et al.*, 1993). Adequate sampling is important for two reasons. First, in undersampled time-frequency representations not all information of the signal is available. After wavelet coding, a signal can be reconstructed by a linear combination

---

[1] A function $f(t)$ has compact support if it is zero outside the interval $T_0<t<T_0+\Delta T$.

of the wavelets, where each wavelet is multiplied by its coefficient (overlap-add procedure). However, undersampled wavelet expansions may lead to far from perfect reconstructions. The second reason is that modifications to undersampled time-frequency representations of sound are affected by the shape of the time-frequency window (Allen and Rabiner, 1977). Interactions between window shape and modification will lead to unwanted byproducts. In this section, different time-frequency tilings will be discussed.

- **Continuous wavelet analysis**

The continuous wavelet analysis is a continuous time-frequency representation: a wavelet coefficient is calculated at every scale and time. Thus, information is available for all times and frequencies. For the continuous wavelet analysis, the sampling density is (theoretically) infinite. As a result, the continuous wavelet analysis is very redundant, and calculation is very time-consuming. Often, it is possible to sample the continuous time-frequency representation and still have essentially all information available, thus being able to reconstruct the original signal. The analysis is still performed in the continuous domain, but it is discrete in the sense that information is available at discrete points in the time-frequency plane. In Chapters 4 and 5 a sampling of the continuous time-frequency plane is used. For adequate sampling in time and frequency, the Nyquist sampling theorem can be used (Allen, 1977; Allen and Rabiner, 1977). For simultaneous time-frequency sampling the Nyquist sampling theorem is applied twice. The sampling interval is based on the temporal and spectral range over which the wavelets are essentially different from zero, meaning that the parts outside this range can be neglected.

- **Discrete wavelet analysis**

The continuous wavelet analysis, whether sampled or not, is performed in the continuous time domain. In contrast, the discrete wavelet analysis is performed in the discrete time domain. Like for the sampled continuous wavelet analysis, this results in information at discrete points in frequency and time. However, the calculation of the discrete wavelet analysis is very different from the calculation of the continuous wavelet transform. The discrete wavelet analysis is calculated by means of successive application of a highpass and a lowpass filter, followed by downsampling by a factor 2. In Fig. 3.5a, such a cascade algorithm is shown. The corresponding scalogram (Fig. 3.3b) shows that for each step, the lower frequency range is divided in 2. This operation improves the spectral resolution

of the lower frequencies. The result is an octave analysis. For each step j (j: integer), going toward lower frequencies, the scale gets a factor of 2 larger. Thus, this operation corresponds to wavelets with scale factor 2^j. However, due to the downsampling operation, the temporal resolution gets a factor of 2 worse in each step. Thus, the temporal shift t_0 is equal to $k \cdot 2^j$ (k: integer), in which k is a counter to cover the whole time range. All information of the signal is available in the resulting wavelet coefficients; the coding is not redundant. This is a very useful property for data compression. A wavelet analysis with Daubechies wavelets can be calculated this way.

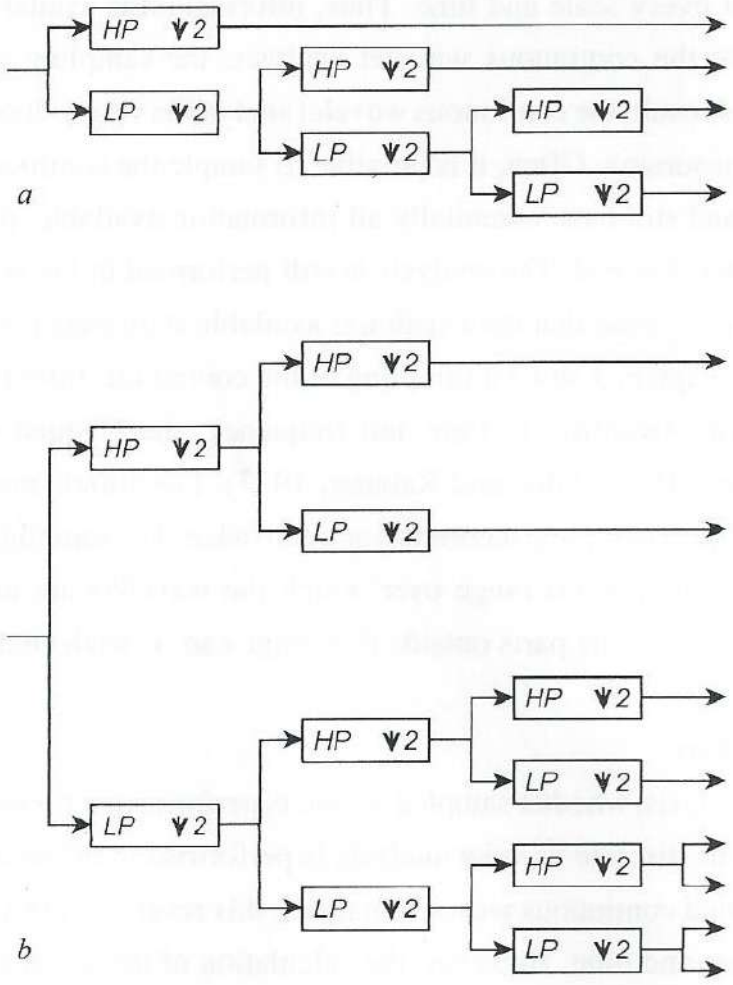


FIG. 3.5. Cascade algorithm of (a) discrete wavelet transform; (b) example of wavelet packet transform. HP: Highpass filter; LP: Lowpass filter; ↓2: downsampling by factor 2.

- **Wavelet packet analysis**

If a spectral resolution of one octave is not good enough, a wavelet packet analysis can be used. The wavelet packet analysis is based on the discrete wavelet analysis. However, in a wavelet packet analysis not only the low-frequency part of the signal is analyzed in a cascade algorithm, but also the high-frequency part (Fig. 3.5b). In this way, an arbitrary frequency split can be obtained (see, e.g., Herley *et al.*, 1993). The price paid is that the joint time-frequency resolution of a wavelet packet analysis is worse than of the discrete wavelet analysis. The analysis functions of a wavelet packet analysis are usually not smooth. Moreover, because of the different succession of highpass and lowpass filters, the resulting analysis functions are not equal to scales and shifts of a mother wavelet.

In this thesis, a Gaussian wavelet was chosen as a mother wavelet. The Gaussian wavelet can only be used in a continuous wavelet transform. Therefore, the Nyquist sampling theorem was used to select an adequate tiling for the Gaussian wavelet. The Gaussian wavelet does not have compact support in time nor in frequency. Therefore, the range between the 25-dB down points was taken as the range over which the window is significantly different from zero (about twice the effective duration and effective bandwidth). This criterion leads to a sampling of one wavelet per three periods of the wavelet carrier frequency along the time axis, and eight wavelets per octave along the frequency axis. It should be noted that, using this time-frequency sampling, the reconstructed signal will not be perfect, because only when the sampling density is infinite, the difference between an original and a reconstructed signal will be zero. However, this sampling density was considered sufficient for its purpose: the difference between original and reconstructed signals was very small and not noticeable to the listeners.

IV. APPLICATIONS

The last ten years, basic wavelet theory has been developed. The territory of applications is less explored. Wavelets can be used for many different applications, from solving partial differential equations to the generation of musical tones. In this section, some applications useful in acoustics will be described: (1) signal analysis, (2) data compression, (3) noise reduction.

The first application of wavelet analysis is in signal analysis. Since wavelet analysis uses short-duration, high frequency wavelets, it is well suited for transient detection (Mallat and Hwang, 1992). It has proven to be useful for high resolution seismic analysis. Since wavelet analysis has important similarities with auditory time-frequency analysis, wavelet analysis is also used to model auditory analysis (Yang *et al.*, 1992; Evangelista, 1993; Irino and Kawahara, 1993; Agerkvist, 1994; Wang and Shamma, 1995; Agerkvist, 1996; this thesis).

Another important application of wavelet analysis is data compression. Using wavelet analysis, it is possible to approximate data with sharp discontinuities by a relatively small number of wavelet coefficients. Especially for image compression, sparse coding by means of wavelet analysis is very successful. In 1993, the US Federal Bureau of Investigation (FBI) adopted a wavelet standard for compression and storage of fingerprints. The 30 million sets of fingerprints are compressed at a ratio of 26:1. Only experts can tell the difference between an original and a compressed fingerprint. Wavelet coding is also used for compression of acoustic signals (Benedetto and Teolis, 1993; Sinha and Tewfik, 1993; Wannamaker and Vrscay, 1997).

A third application of wavelet analysis is the de-noising of noisy data. In a de-noising algorithm, the wavelet coefficients are subjected to a nonlinear threshold operation. In a hard-thresholding operation, all coefficients with a modulus less than a certain value are set to zero. In a soft-thresholding operation, coefficients with modulus less than a certain value are attenuated. The idea is that coefficients with few energy probably do not contain the important information of the signal, but noise. By making these zero or by attenuation, this noise may be reduced. With respect to sounds, wavelet analysis has been used for speech enhancement (Drake *et al.*, 1993; Pintér, 1996; Whitmal *et al.*, 1996;

Nishimura *et al.*, 1998). Another example of de-noising is the removal of scratch noise in old recordings (Montresor *et al.*, 1990). Then, the zooming-in property of wavelet analysis is successful in the detection of the scratches (edge detection).

V. AUDITORY WAVELET CODING

In this chapter, the similarities between wavelet coding and peripheral auditory coding were discussed. By choosing a Gaussian mother wavelet with a bandwidth of ¼ octave, the joint time-frequency resolution of the wavelet coding was roughly matched to that of the auditory system. Given this mother wavelet, an adequate time-frequency tiling is eight wavelets per octave along the frequency axis, and one wavelet every three periods along the time axis. This wavelet coding can be seen as a perceptually relevant wavelet coding, and is taken as a model of normal peripheral auditory coding. In the following chapters, a distortion of the wavelet coding will be used to model distorted peripheral auditory coding in hearing-impaired listeners.

REFERENCES

Agerkvist, F. T. (**1994**). "Time-frequency analysis and auditory models," doctoral dissertation (ISSN 0105-3027), Technical University of Denmark, Denmark.

Agerkvist, F. T. (**1996**). "A time-frequency auditory model using wavelet packets," J. Audio Engineering Society **44**, 37–50.

Allen, J. B. (1977). "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," IEEE Trans. Acoust. Speech Signal Process. **25**, 235–238.

Allen, J. B., and Rabiner, L. R. (**1977**). "A unified approach to short-time Fourier analysis and synthesis," Proc. of the IEEE **65**, 1558–1564.

Benedetto, J. J., and Teolis, A. (**1993**). "A wavelet auditory model and data compression," Applied and computational harmonic analysis 1, 3–28.

Drake, L. A., Rutledge, J. C., and Cohen, J. (**1993**). "Wavelet analysis and recruitment of loudness compensation," IEEE Trans. Signal Processing **41**, 3306–3312.

Evangelista, G. (**1993**). "Pitch-synchronous wavelet representation of speech and music signals," IEEE Trans. Signal Processing **41**, 3313–3330.

Gabor, D. (**1947**). "Acoustical quanta and the theory of hearing," Nature (London) **159**, 591–594.

Herley, C., Kovačević, J., Ramchandran, K., and Vetterli, M. (**1993**). "Tilings of the time-frequency plane: construction of arbitrary orthogonal bases and fast tiling algorithms," IEEE Trans. Signal Processing **41**, 3341–3359.

Irino, T., and Kawahara, H. (**1993**). "Signal reconstruction from modified auditory wavelet transform," IEEE Trans. on signal processing **41**, 3549–3554.

Irino, T., and Patterson, R. D. (**1996**). "Temporal asymmetry in the auditory system," J. Acoust. Soc. Am. **99**, 2316–2331.

Irino, T., and Patterson, R. D. (**1997**). "A time-domain, level-dependent auditory filter - the gammachirp," J. Acoust. Soc. Am. **101**, 412–419.

Landau, H. J., and Pollak, H. O. (**1961**). "Prolate spheroidal wave functions, Fourier analysis and uncertainty – II," The Bell System Technical Journal **40**, 65–84.

Mallat, S., and Hwang, W. L. (**1992**). "Singularity detection and processing with wavelets," IEEE Trans. on Information Theory, Special Issue on Wavelet Transforms and Multiresolution Signal Analysis **38**, 617–643.

Montresor, S., Valiere, J. C., Allard, J. F., and Baudry, M. (**1990**). "The restoration of old recordings by means of digital techniques," A. E. S. Montreux, Preprint 2915 (G4).

Nishimura, R., Asano, F., Suzuki, Y., and Sone, T. (**1998**). "Speech enhancement using spectral subtraction with wavelet transform," Electronic and Communications in Japan Part 3, Vol. **81**, No. 1, 24–31.

Pintér, I. (**1996**). "Perceptual wavelet-representation of speech signals and its application to speech enhancement," Computer speech and language **10**, 1–22.

Rioul, O., and Flandrin, P. (**1992**). "Time-scale energy distributions: A general class extending wavelet transforms," IEEE Trans. Signal Process. **40**, 1746–1757.

Rioul, O., and Vetterli, M. (**1991**). "Wavelets and signal processing," IEEE Signal Proc. Mag. October, 14–38.

Scharf, B. (**1970**). "Critical bands," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. 1, pp. 157–202.

Sinha, D., and Tewfik, A. H. (**1993**). "Low bit rate transparent audio compression using adapted wavelets," IEEE Trans. on signal processing **41**, 3463–3479.

Strang, G., and Nguyen, T. (**1996**). "Wavelets and filter banks," Wellesley-Cambridge Press (Wellesley, USA).

Vetterli, M., and Kovačević, J. (**1995**). "Wavelets and subband coding," Prentice Hall (New York).

Wang, K., and Shamma, S. A. (**1995**). "Spectral shape analysis in the central auditory system," IEEE Trans. on Speech and Audio Processing **3**, 382–395.

Wannamaker, R. A., and Vrscay, E. R. (**1997**). "Fractal wavelet compression of audio signals," J. Audio Eng. Soc. **45**, 540–553.

Whitmal, N. A., Rutledge, J. C., and Cohen, J. (**1996**). "Reducing correlated noise in digital hearing aids," IEEE Engineering in Medicine and Biology September/October, 88–96.

Yang, X., Wang, K., and Shamma, S. A. (**1992**). "Auditory representation of acoustic signals," IEEE Trans. on Information Theory **38**, 824–839.

4 The effect of intensity perturbations on speech intelligibility for normal-hearing and hearing-impaired listeners

Hearing-impaired listeners are known to suffer from reduced speech intelligibility in noise, even if sounds are above their hearing thresholds. This study examined the possible contribution of reduced acuity of intensity coding to this problem. The "distortion-sensitivity model" was used: the effect of reduced acuity of auditory intensity coding on intelligibility was mimicked by an artificial distortion of the speech intensity coding, and the sensitivity to this distortion for hearing-impaired listeners was compared with that for normal-hearing listeners. Stimuli (speech plus noise) were wavelet coded using a Gaussian wavelet (1/4 octave bandwidth). The intensity coding was distorted by multiplying the modulus of each wavelet coefficient by a random factor. Speech-reception thresholds (SRTs) were measured for various degrees of intensity perturbation. Hearing-impaired listeners were classified as suffering from suprathreshold deficits if intelligibility of undistorted speech was worse than predicted from audibility by the Speech Intelligibility Index model (ANSI, 1997). Hearing-impaired listeners without suprathreshold deficits were as sensitive to the intensity distortion as the normal-hearing listeners. Hearing-impaired listeners with suprathreshold deficits appeared to be less sensitive. Results indicate that reduced acuity of auditory intensity coding may be a factor underlying reduced speech intelligibility for the hearing impaired.

Submitted to the Journal of the Acoustical Society of America

INTRODUCTION

Speech recognition (or intelligibility) in noisy environments is a problem for many hearing-impaired listeners. This problem may result from inaudibility of part of the speech spectrum. However, even if sounds are above hearing thresholds over the whole frequency range, some hearing-impaired listeners still have problems perceiving speech in noise. Their speech processing is not as good as that of normal-hearing listeners due to suprathreshold deficits (Moore, 1996; Noordhoek *et al.*, in press). Examples of suprathreshold deficits are reduced spectral resolution (frequency selectivity), reduced temporal resolution, impaired frequency discrimination, or impaired loudness perception.

This study examines a deficit related to impaired loudness perception, i.e., reduced intensity coding. Reduced intensity coding may be thought of as a less accurate intensity representation in the auditory periphery. This may be due, for example, to a loss of auditory nerve fibers, resulting in a more noisy intensity coding. Reduced intensity coding may lead to higher just-noticeable differences (jnd's) in intensity or less jnd's. A few studies suggest that intensity coding may be disrupted for some listeners with cochlear damage (Florentine *et al.*, 1993; Buus *et al.*, 1995; Moore, 1995). Reduced intensity coding acuity is likely to affect speech intelligibility. However, the literature does not report any attempts to relate intensity coding to speech intelligibility.

The main question in this study is whether poor auditory intensity coding is at least partly responsible for the observed poor speech intelligibility in noise by hearing-impaired listeners. This is examined by introducing an artificial distortion in the intensities of speech. The distortion simulates the effect of reduced acuity of auditory intensity coding on speech perception. Speech-reception thresholds (SRTs) for various degrees of the applied artificial distortion are compared for normal-hearing and hearing-impaired listeners in order to clarify the contribution of reduced auditory intensity coding acuity to impaired speech intelligibility. This type of experiment may be called a "distortion-sensitivity approach" (Houtgast, 1995).

Under the distortion-sensitivity model, a specific type of distortion is applied to combined speech and noise stimulus. Intelligibility tests are administered in order to determine whether the artificial distortion is or is not related to the suprathreshold deficit

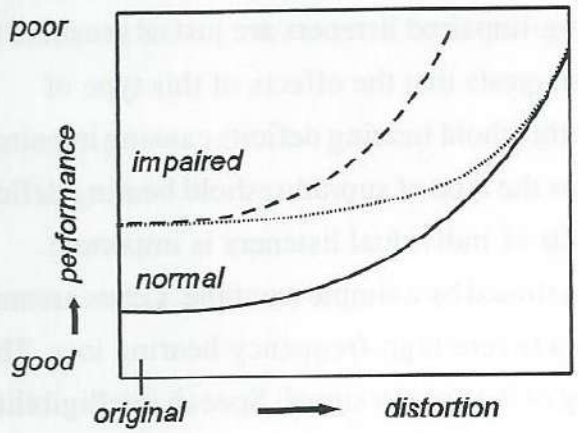


FIG. 4.1. Illustration of the distortion-sensitivity model. Performance measured as a function of the degree of distortion of hearing-impaired listeners is compared with that of normal-hearing listeners (solid line). The possible outcome of such an experiment is "convergence" (dotted line) or "no convergence" (dashed line). See the text for further explanation.

of hearing-impaired listeners. Therefore, intelligibility is measured as a function of the degree of the distortion, and sensitivity to the distortion is compared for normal-hearing and hearing-impaired listeners. Fig. 4.1 is a schematic illustration of the possible outcomes of such an experiment.

In the undistorted condition, using the original, unmodified speech, hearing-impaired listeners usually recognize speech more poorly than normal-hearing listeners. When comparing speech intelligibility by normal-hearing and hearing-impaired listeners as a function of the degree of distortion, essentially, two different trends may be hypothesized. First, performance of the normal-hearing and hearing-impaired listeners converges (dotted line). Second, performance of the normal-hearing and hearing-impaired listeners does not converge (dashed line).

In the convergence case, the performance difference between normal-hearing and hearing-impaired listeners becomes smaller as a function of the distortion level. For high levels of distortion, performance becomes essentially equal. Thus, hearing-impaired listeners are less sensitive to the distortion than normal-hearing listeners. In statistical terms, this is an interaction between listener groups and level of distortion or, stated differently, between hearing deficit and distortion. In terms of interpretation, the effect of the artificial distortion is smaller for hearing-impaired listeners because the hearing deficit already affects the speech processing in a similar way. Thus, the specific type of artificial distortion for which convergence is observed, hints at the suprathreshold deficit causing the speech intelligibility problems of the hearing impaired.

In the no-convergence case, performance by normal-hearing and hearing-impaired listeners does not come close together. Hearing-impaired listeners are just as sensitive to distortion as normal-hearing listeners. This suggests that the effects of this type of artificial distortion are not related to the suprathreshold hearing deficits causing impaired speech intelligibility. It should be noted that, as the type of suprathreshold hearing deficit may be listener dependent, studying the results of individual listeners is important.

The distortion-sensitivity model can be illustrated by a simple example. Let us assume that a hearing-impaired listener suffers from a severe high-frequency hearing loss. The applied artificial distortion is lowpass filtering of the speech signal. Speech intelligibility is measured as a function of the cutoff frequency of the lowpass filter. Lowpass filtering reduces the speech intelligibility. Compared with normal-hearing listeners, the hearing-impaired listener is less sensitive to the lowpass filtering. This is because the high frequencies in the broadband signal are not perceived anyway. Convergence of the performance of normal-hearing and hearing-impaired listeners suggests that lowpass filtering relates to the problem experienced by the hearing-impaired listener, i.e., the listener misses some part of the high-frequency spectrum.

In this study, *artificial* distortion of the intensity coding tries to mimic poor *auditory* intensity coding. To simulate poor intensity coding, a model of 'normal' auditory intensity analysis is required. Auditory analysis is a spectro-temporal analysis. This is modeled by wavelet decomposition. Wavelet analysis is used for mimicking auditory time-frequency analysis because of its logarithmic frequency scale (see, e.g., Rioul and Vetterli, 1991). In Chapters 2 and 3 it was shown that auditory spectral and temporal resolution are roughly matched by using a Gaussian-shaped mother wavelet (prototype analysis function) with a bandwidth of 1/4 octave. Using this perceptually relevant time-frequency analysis, specific manipulations of the wavelet coefficients may be used to simulate specific changes in auditory coding. Therefore, a reduced acuity in auditory intensity coding may be simulated by introducing random perturbation in the intensity of the wavelet coefficients.

In summary, the aim of this study is to investigate if reduced speech intelligibility by hearing-impaired listeners may be explained by reduced intensity coding. This question is addressed by a "distortion-sensitivity model" in which an artificial distortion of the intensities in a speech-plus-noise stimulus between wavelet decomposition and recomposition is applied. Intelligibility is measured as a function of the degree of

distortion, and the sensitivity of hearing-impaired and normal-hearing listeners is compared. The rationale behind the distortion-sensitivity model is that when a hearing-impaired listener is less sensitive to the intensity distortion than normal-hearing listeners this may indicate that poor auditory intensity coding is causing part of the speech intelligibility problems.

I. METHOD

A. Distortion of wavelet coded intensities

In this study, intensity coding of sound is distorted to mimic the effects of poor auditory intensity coding. By means of the Speech-reception threshold test (SRT; for an explanation, see Sec. 4.I D2), speech intelligibility of sentences is measured as a function of the degree of applied artificial intensity distortion. In order to simulate auditory intensity coding, a perceptually relevant spectro-temporal analysis method has been developed.

To model auditory spectro-temporal coding, sounds were described in the time-frequency domain by means of a wavelet transform. Compared with the short-time Fourier transform, the wavelet transform matches auditory system coding more closely because it uses a logarithmic frequency scale (e.g., Rioul and Vetterli, 1991). In this study, the criterion for the choice of the mother wavelet is its spectral (and temporal) resolution. Results of Chapter 2 suggest that a Gaussian-windowed sinusoid with a shape factor between 0.15 and 0.3 roughly matches the auditory time-frequency window. Therefore, as the prototype analysis function, a complex sinusoidal carrier with a Gaussian envelope was chosen. This Gaussian wavelet is described by

$$s(t) = \sqrt{\alpha f_0}\, \exp(i2\pi f_0 t)\, \exp(-\pi(\alpha f_0 t)^2) \quad , \tag{4.1}$$

in which f_0 is the carrier frequency, α is the shape factor, and $\sqrt{\alpha f_0}$ normalizes the energy of the analysis function. This time-frequency window has an effective bandwidth of $\Delta_f = \alpha f_0$ and an effective duration of $\Delta_t = 1/(\alpha f_0)$ (see Gabor, 1947). The shape factor α

was set to 0.1735. Thus, the effective bandwidth of the analysis function was ¼ octave [about the auditory critical band (see Scharf, 1970)]. As a result, the effective duration of the time-frequency window is 5.76 ms at 1 kHz (1.44 ms at 4 kHz). The effective number of sinusoidal periods contained within the Gaussian envelope equals 5.8 (i.e., $1/\alpha$).

This Gaussian wavelet is used to construct a wavelet decomposition that covers the time-frequency plane. Shifts of this prototype function cover the temporal domain; scales of the prototype function cover the spectral domain. The scaling is controlled by varying the carrier frequency f_0. For simultaneous sampling in time and frequency the Nyquist sampling theorem was used (see Allen, 1977; Allen and Rabiner, 1977). This theorem is based on the bandwidth and duration of the analysis function. Because the Gaussian wavelet does not have compact support[1] in time nor in frequency, the 25-dB down points were taken as an estimate of the upper limit of bandwidth and duration of the analysis functions. This leads to a sampling of one wavelet per three periods of the wavelet carrier frequency along the time axis, and eight wavelets per octave along the frequency axis. The theoretical number of complex coefficients needed to describe the signal is about 2 per input sample (see Allen, 1977). In this study, the information of the signals was limited to the frequency range from 250 to 4000 Hz. As a result, the number of coefficients computed per input sample could be limited to about unity. Thus, one second of speech (sampling frequency: 15625 Hz) was described by $16*10^3$ complex wavelet coefficients in which no information below 250 Hz and above 4 kHz was preserved.

Using these coefficients, sounds can be reconstructed by an overlap-add procedure. Theoretically, the reconstruction is not perfect. However, using the above described time-frequency tiling, differences between the original signal and the reconstructed signal are very small and not noticeable to a listener.

After the wavelet analysis, the modulus of each wavelet coefficient was perturbed to mimic the effect of a reduced accuracy in intensity coding. This was achieved by multiplying each individual complex wavelet coefficient by a random factor. As a result, silence will still be silence after perturbation. The random perturbation factor ε (in dB) was chosen from a uniform distribution with zero mean and boundaries *-Pmax* and *+Pmax*. Thus, the modulus of each individual coefficient was multiplied by a different

---

[1]A function *f(t)* has compact support if it is zero outside the interval $T_0<t<T_0+\Delta T$.

random factor $10^{\varepsilon/20}$. After perturbation, the energy contained in each frequency band over the whole sentence was scaled to equal the original energy in this band.

The perturbation of the intensity coding was applied to the combined speech and noise signal. This probably simulates impaired auditory processing more realistically than a procedure in which speech and noise are processed separately and then combined.

B. Subjects

Twenty-five sensorineurally hearing-impaired listeners participated in the experiment. They were all native Dutch speakers, aged 24 to 70 years with a mean age of 41 years. Their intelligibility scores for monosyllabic words in quiet were at least 75% correct. Thresholds in the better-hearing ears averaged over 0.5, 1, and 2 kHz (the pure-tone average, or PTA) ranged from 7 to 58 dB HL, with a mean PTA of 38 dB HL. The pure-tone, air-conduction thresholds in the better-hearing ears were at least 30 dB HL at one or more frequencies between 250 and 4000 Hz.

Twenty-two normal-hearing listeners (aged 19 to 29 years with a mean age of 22 years) served as a control group. All were native Dutch speakers. Pure-tone air-conduction thresholds of the normal-hearing listeners did not exceed 15 dB HL at any octave frequency from 250 to 4000 Hz.

C. Stimuli and apparatus

Speech material consisted of lists of 13 everyday Dutch sentences of eight to nine syllables read by a female or a male speaker (Plomp and Mimpen, 1979; Smoorenburg, 1992). The masking noise was spectrally shaped for each speaker individually according to the long-term average spectrum of all sentences.

Signals were generated by TDT (Tucker Davis Technologies System II) hardware. Stimuli were presented in the middle of the dynamic range of each listener by frequency shaping using a programmable filter (TDT PF1). The stimuli were presented monaurally through Sony MDR-V900 headphones. To avoid the risk of cross-hearing, the listener's better-hearing ear was tested. For calibration, noise levels were measured on a Brüel & Kjær type 4152 artificial ear with a flat-plate adapter. The entire experiment was

controlled via a personal computer. Subjects were tested individually in a soundproof room.

D. Procedures

First, the dynamic range of each listener was determined. Then, speech intelligibility was measured, in which combined speech and noise were presented in the middle of the dynamic range. These tests are described below. To familiarize the subjects with the procedure, a training session preceded data collection. All conditions were measured twice to determine test-retest reliability. An essential part of the distortion-sensitivity model is the comparison of the performance of individual hearing-impaired listeners with that of normal-hearing listeners. Therefore, for all listeners, the same order of conditions was used. In addition, a different but fixed sentence list was used in each condition.

- **Dynamic Range**

The dynamic range of each listener was estimated by measuring the hearing threshold and the uncomfortable loudness level (UCL) for narrow bands of noise. The UCL was corrected for broadband stimulation, as described below.

Thresholds and UCLs were measured with 1/3-octave noise bands with center frequencies at 250, 500, 1000, 2000, and 4000 Hz. Hearing thresholds were measured using a Békésy tracking procedure (300-ms noise bursts; repetition rate 2.5 Hz; step size 1 dB). The measurement was ended after eleven reversals. The average of all but the first reversal level was taken as the hearing threshold. Narrow-band UCLs were measured with noise bursts presented increasing in level by 3 dB for each presentation (300 ms noise burst; repetition rate 1.4 Hz). Listeners were asked to push a button when the noise bursts became uncomfortably loud. Then, the level of the noise burst was immediately diminished by a random amount between 21 and 30 dB, and the ascending procedure was repeated until six responses were obtained. The average of the levels at which the button was pushed was taken as the narrow-band UCL.

To correct the UCL for broadband stimulation, a 4-second broadband noise burst was presented, spectrally shaped according to the narrow-band UCLs and starting 40 dB below the narrow-band UCLs. The level of the broadband noise burst was gradually increased in steps of 5 dB. After each presentation the listener was asked whether the

signal was experienced as uncomfortably loud. If this was the case, the corresponding level was taken as the broadband UCL.

- **Speech intelligibility**

Speech-reception threshold in noise for an adapted spectrum (SRTa)

The speech-reception threshold (SRT, Plomp and Mimpen, 1979) was used to measure speech intelligibility. The SRT in noise is defined as the signal-to-noise ratio (SNR) at which 50% of sentences are reproduced correctly. The speech level is varied in an adaptive, up-down procedure with a step size of 2 dB. Speech and noise are adapted to fit in the dynamic range of individual listeners. The adapted speech-reception threshold is called SRTa. In the SRTa tests in this study, all stimuli were bandpass filtered from 250 to 4000 Hz. The SRTa was measured as a function of intensity perturbation.

The aim of this study is to assess the effect of a reduced *auditory* intensity coding resulting from *artificial* perturbations of the intensity coding of the speech-plus-noise stimulus. Because of the applied intensity perturbations, the auditory system is not provided with accurate intensity information. However, the applied intensity perturbations also introduce spectro-temporal fluctuations. To study the effects of distorted intensity coding, it is important to ensure that spectro-temporal effects do not dominate the speech intelligibility of hearing-impaired listeners. Therefore, in the present study speech intelligibility was measured for intensity perturbations that only slightly affect performance. Preliminary data were collected to determine the appropriate range of intensity perturbations to apply.

The SRTa was measured as a function of the degree of intensity perturbation (0, 10, 20, 30, and 40 dB) for 10 normal-hearing listeners. Fig. 4.2 presents the results. Mean data for the normal-hearing listeners are indicated by open symbols. Error bars indicate the standard error of the mean. A typical example of the performance of a hearing-impaired listener is indicated by filled symbols. For the normal-hearing listeners at 10 dB, the SRTa is slightly affected (difference with no perturbation: 1.3 dB). For more severe perturbations, the SRTa increases almost linearly with perturbation, ranging from -0.9 dB when *Pmax* is 10 dB to 6.6 dB when *Pmax* is 40 dB. The hearing-impaired listener appears to be hardly affected by the 10 dB intensity perturbation; the decrease in performance compared with the reference condition (no perturbation) is only 0.3 dB. However, for larger degrees of intensity perturbations, speech intelligibility deteriorates

more quickly than observed for the normal-hearing listeners. Linear regression lines were fitted through the SRTa data at 10, 20, 30, and 40 dB of intensity perturbation. The slope of the hearing-impaired listener was steeper than the 95-percent upper boundary of the slopes of the normal-hearing listeners.

These data suggest that severe degrees of intensity perturbation affected this hearing-impaired listener more than the normal-hearing listeners. This may be explained by the effect of the perturbations on loudness perception. Since the dynamic range of the hearing-impaired listener was markedly smaller than that of the normal-hearing listeners, the same intensity perturbation did not result in the same loudness perturbation. This hearing-impaired listener was probably subjected to higher degrees of loudness perturbation than the normal-hearing listeners. Another cause may be the spectro-temporal fluctuations introduced by the artificial intensity distortion. These fluctuations in the combined speech and noise may result in additional masking in the temporal domain, i.e., forward and backward masking, and in the spectral domain, i.e., upward and downward spread of masking. Hearing-impaired listeners are known to suffer from excessive masking. [For review, see Moore (1995).] For large amounts of intensity perturbations these unwanted spectro-temporal byproducts may even dominate the speech intelligibility of hearing-impaired listeners, causing performances for normal-hearing and hearing-impaired listeners to diverge.

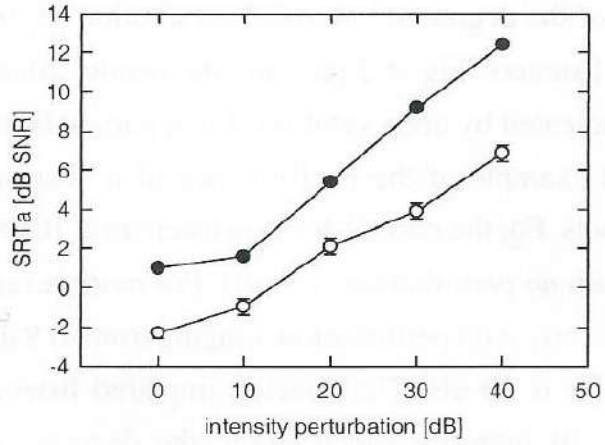


FIG. 4.2. Average SRTa as a function of intensity perturbation for ten normal-hearing listeners (open symbols). The error bars represent the standard error of the mean. Also, a typical example of the performance for hearing-impaired listeners is plotted (filled symbols: data for one hearing-impaired listener).

To recapitulate, a small but consistent effect on speech intelligibility was observed for intensity perturbations of 10 dB. To avoid the risk of spectro-temporal effects of the intensity distortion algorithm, the range 0–10 dB was measured in the speech intelligibility experiment. As a measure for the sensitivity to the intensity distortion, the SRTa at 10 dB minus the SRTa at 0 dB is used.

Speech-Reception Bandwidth Threshold (SRBT)

To classify the hearing-impaired listeners into a group "with" and "without" suprathreshold deficits, the Speech-Reception *Bandwidth* Threshold (SR*B*T) was measured. The SR*B*T is a measure of speech intelligibility introduced by Noordhoek *et al.* (1999). The SR*B*T is highly sensitive for suprathreshold deficits, as is shown in a recent study of Noordhoek *et al.* (in press).

The SR*B*T procedure is similar to the SRT procedure, except that the bandwidth (center frequency 1 kHz) of speech sounds is varied instead of their levels when estimating the 50% intelligibility threshold. Complementary bandstop noise is added to the bandpass-filtered speech. Both speech and noise are presented in the middle of the listener's dynamic range.

E. Speech intelligibility index

As a measure for the quality of speech processing, the SRTa and SR*B*T data were converted to a Speech Intelligibility Index. The Speech Intelligibility Index (SII) (ANSI, 1997) is a physical measure of how much information of the speech is available to the listener. The SII model accounts for hearing threshold, self-masking in speech, upward spread of masking and level distortion at high presentation levels. To calculate the SII, speech spectra, noise spectra and hearing thresholds must be known. Therefore, sound pressure levels of speech and noise (divided in 1/3-octave bands) were measured with the headphone positioned on a Brüel & Kjær type 4152 artificial ear with a flat-plate coupler. These levels were converted to equivalent free-field levels.

II. RESULTS AND DISCUSSION

A. Suprathreshold deficits

In the speech intelligibility tests (SRTa and SR*B*T) sounds were spectrally shaped to fit in the dynamic range of individual listeners. A comparison of the results for normal-hearing and hearing-impaired listeners provides insight into the speech intelligibility performance of the hearing impaired when sounds are presented above hearing threshold.

For the normal-hearing listeners, the average SRTa was -2.1 dB (standard deviation 0.9 dB); for the hearing-impaired listeners, the SRTa ranged from -2.0 dB to 6.8 dB, with an average of 0.4 dB. The individual standard error (test-retest) averaged over all listeners was 1.1 dB.

For the normal-hearing listeners, the average SR*B*T was 1.44 octave (standard deviation 0.18 octave); for the hearing-impaired listeners, the SR*B*T ranged from 1.25 to 3.49 octave, with an average of 1.94 octave. The individual standard error (test-retest) averaged over all listeners was 0.16 octave.

The upper limit of the one-tailed 95% confidence interval of the data for the normal-hearing listeners was used to distinguish the hearing-impaired listeners with difficulty of listening in noise. Relative to this boundary, the SRTa was elevated for 15 of the 25 hearing-impaired listeners; the SR*B*T was elevated for 13 of the 25 hearing-impaired listeners. This indicates that a substantial number of the hearing-impaired listeners has problems recognizing speech in noise, even if sounds are presented in the middle of the dynamic range of the listeners.

Speech intelligibility problems may be due to suprathreshold deficits. However, other possible explanations are inaudibility of part of the speech spectrum (if the dynamic range of a listener is very small) or high presentation levels causing extra upward spread of masking and level distortion. Therefore, to investigate the effect of suprathreshold deficits on speech intelligibility, individual SRTa and the SR*B*T data were converted into SII units. An elevation of the SII-values of a hearing-impaired listener compared with that of the normal-hearing listeners indicates the presence of suprathreshold deficits. The higher the SII, the more serious the speech processing deficits. Fig. 4.3 shows the individual SII

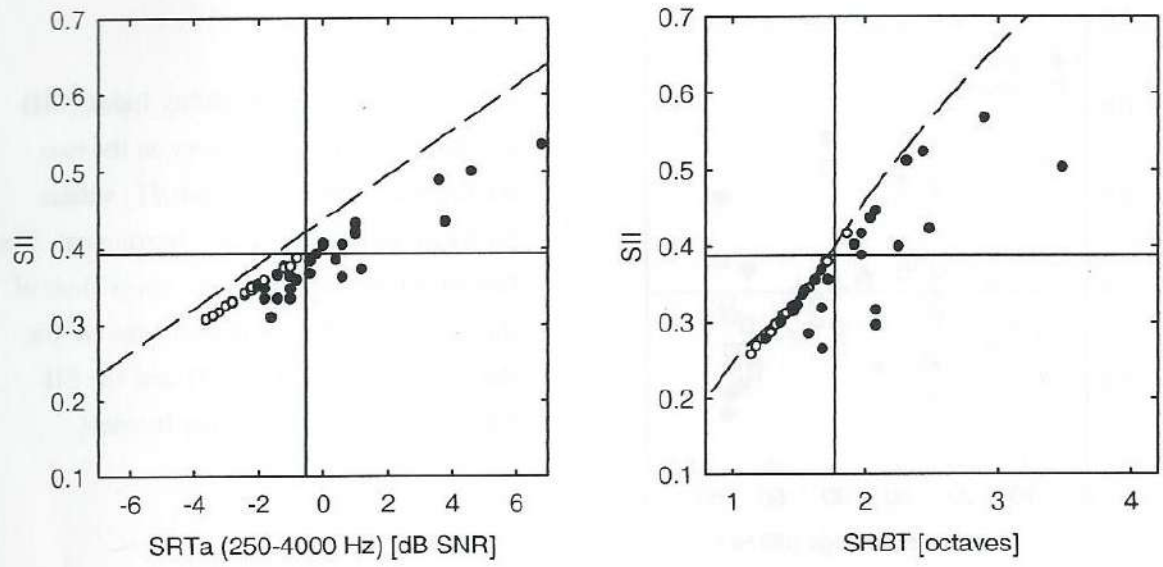


FIG. 4.3. Speech Intelligibility Index (SII) versus SRTa and SR*B*T for normal-hearing listeners (open circles) and hearing-impaired listeners (filled circles). Solid lines represent the upper boundaries of the one-tailed 95% confidence intervals for normal-hearing listeners. Dashed lines represent the maximum SII when the audibility of the speech is not influenced by the hearing threshold, upward spread of masking, and level distortion.

values of the SRTa and SR*B*T test for the normal-hearing listeners (open circles) and the hearing-impaired listeners (filled circles). The SII values are plotted as a function of the individual results on the two speech intelligibility tests. The upper limit of the one-tailed 95% confidence interval of the SII's of the normal-hearing listeners is chosen as the boundary between normal and elevated SII. This is indicated by a horizontal solid line. The boundary between normal and elevated SRTa or SR*B*T is indicated by a vertical solid line. The dashed lines in Fig. 4.3 represent the maximum SII value when the audibility of the speech is not influenced by the hearing threshold, upward spread of masking, and level distortion. 10 of the 25 hearing-impaired listeners have a higher than normal SII-SRTa and 11 have a higher than normal SII-SR*B*T; of the latter group, 7 also have a higher than normal SII-SRTa. These results show that a substantial number of hearing-impaired listeners have speech intelligibility problems because of suprathreshold deficits. In Sec. 4.II B hearing-impaired listeners are divided into groups *with* and *without* suprathreshold deficits.

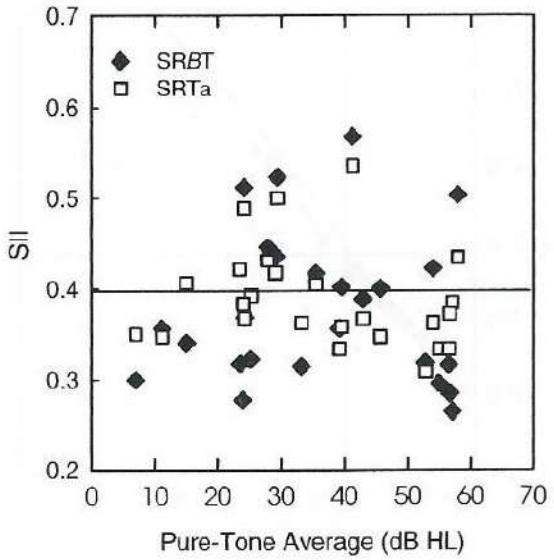


FIG. 4.4. Speech Intelligibility Index (SII) for hearing-impaired listeners on the two intelligibility tests (SRTa, SR*B*T) versus pure tone average for better hearing ear. The horizontal line represents the upper limit of the one-tailed 95% confidence interval for the SII's, both the SII-SR*B*T and the SII-SRTa, of the normal-hearing listeners.

The relation between the occurrence of suprathreshold deficits and hearing loss is illustrated in Fig. 4.4. SII-SR*B*T and SII-SRTa for the hearing-impaired listeners are plotted as a function of PTA. The horizontal line is the 95% confidence limit of the SII's (SII-SR*B*T and SII-SRTa combined) for the normal-hearing listeners. Figure 4.4 shows no correlation between hearing loss and SII. This indicates that some hearing-impaired listeners with only a mild hearing loss experienced hampered speech perception due to suprathreshold deficits. In contrast, some hearing-impaired listeners with a severe hearing loss did not suffer from suprathreshold deficits. This finding agrees with the results of Noordhoek *et al.* (in press).

B. The distortion-sensitivity model

The distortion-sensitivity model compares speech intelligibility as a function of the degree of distortion for normal-hearing and hearing-impaired listeners. The aim is to determine whether artificial distortion relates to a suprathreshold deficit causing impaired speech perception. The hypotheses underlying this model were schematically illustrated in Fig. 4.1. In Fig. 4.5 the results of this study are depicted.

The average results of the normal-hearing listeners are represented by the open circles. The hearing-impaired listeners are divided into two groups: (1) *without* suprathreshold deficits and (2) *with* suprathreshold deficits. The division is based on the SII-SR*B*T

because the SII-SR*B*T is independent of the values plotted in Fig. 4.5 and, as was mentioned in Sec. 4.I D2, the SR*B*T test is highly sensitive to suprathreshold deficits. This resulted in a group of 14 listeners *without* suprathreshold deficits, of which the average SRTa-values are represented by downward pointing triangles, and a group of 11 listeners *with* suprathreshold deficits, of which the average scores are represented by upward pointing triangles. Not all listeners were tested at 5 dB of intensity perturbation. Data points in this condition are for 12 normal-hearing listeners, and for 5 hearing-impaired listeners *with* suprathreshold deficits and 8 *without*. The error bars represent the standard error of the mean.

Compared with normal-hearing listeners, hearing-impaired listeners *without* suprathreshold deficits show SRTa measures that are shifted upwards by 1 dB. No convergence of the data is observed. A student t-test with unequal variances comparing the 'sensitivity to the distortion' (SRTa at 10 dB minus SRTa at 0 dB) of normal-hearing and hearing-impaired listeners did not show significant convergence either. This group of hearing-impaired listeners is just as sensitive to the distortion as the normal-hearing listeners.

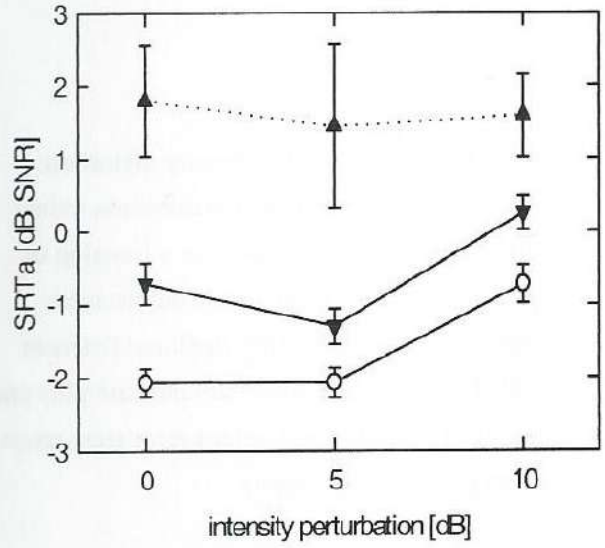


FIG. 4.5. SRTa as a function of intensity perturbation for normal-hearing listeners (open circles) and hearing-impaired listeners. Hearing-impaired listeners are divided into two groups: *with* (upward pointing triangles) and *without* (downward pointing triangles) suprathreshold deficits. Data points in conditions 0 and 10 dB intensity perturbation are for 22 normal-hearing listeners, and for 11 hearing-impaired listeners *with* suprathreshold deficits and 14 *without*. Data points in condition 5 dB intensity perturbation are for 12 normal-hearing listeners, and for 5 hearing-impaired listeners *with* suprathreshold deficits and 8 *without*. The error bars show the standard error of the mean.

The performance of the hearing-impaired listeners *with* suprathreshold deficits does converge toward the performance of the normal-hearing listeners for increasing amounts of perturbation. This implies that the hearing-impaired listeners with suprathreshold deficits are less sensitive to the distortion than the normal-hearing listeners. A student t-test with unequal variances confirmed this ($p<0.05$).

Thus, the hearing-impaired listeners *without* suprathreshold deficits are as sensitive to the intensity perturbations as the normal-hearing listeners. This is not surprising since the SII model shows that their speech intelligibility problems can be explained solely on basis of audibility. Their suprathreshold speech processing is as good as that of normal-hearing listeners. However, the hearing-impaired listeners *with* suprathreshold deficits are less sensitive to the intensity distortion. As already mentioned in Sec. 4.I D2, the same degree of intensity perturbation will result in a larger degree of loudness perturbation for hearing-impaired listeners than for normal-hearing listeners, because the hearing-impaired listeners have a smaller dynamic range than the normal-hearing listeners. However, the conversion of the intensity factor to a loudness perturbation factor for each listener will result in a more pronounced convergence of performance for normal-hearing and hearing-impaired listeners. In conclusion, along the lines of the distortion-sensitivity model, the results suggests that the artificial intensity distortion is related to the suprathreshold speech-processing problems of hearing-impaired listeners.

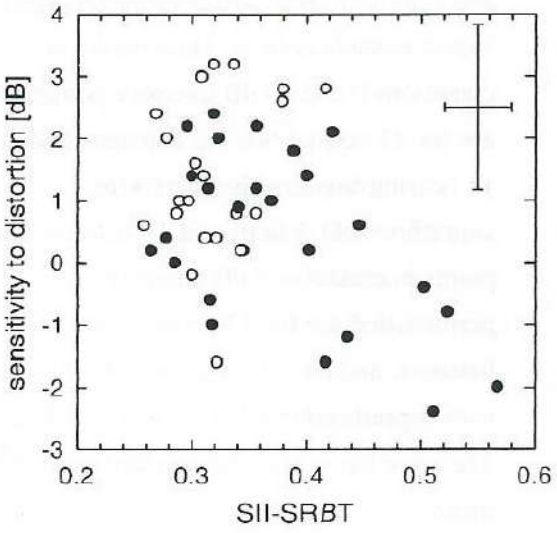


FIG. 4.6. Sensitivity to intensity distortion (SRTa at 10 dB intensity perturbation minus SRTa without perturbation) as a function of the SII-SR*B*T for normal-hearing listeners (open circles) and hearing-impaired listeners (filled circles). The error bars indicate plus and minus the individual standard error (test-retest) averaged over all listeners.

The division of the hearing-impaired listeners into two groups showed that hearing-impaired listeners *with* suprathreshold deficits are less sensitive to the intensity distortion than normal-hearing listeners, whereas the group *without* is just as sensitive as the normal-hearing listeners. To explore this relation between suprathreshold deficits and distortion sensitivity further, it is interesting to look at the individual results. As mentioned in Sec. 4.I D2, the difference in SRTa between 10 dB and 0 dB intensity perturbation was taken as a measure of the individual sensitivity to the distortion. In Fig. 4.6, this sensitivity is plotted as a function of the SII-SR*B*T. Open symbols represent the data for normal-hearing listeners, the filled symbols those for hearing-impaired listeners. The individual standard error of the SII-SR*B*T (test and retest) averaged over all listeners was 0.029; the individual standard error of the sensitivity to the distortion (test and retest) averaged over all listeners was 1.3 dB. The error bars indicate plus and minus one individual standard error. For some listeners, sensitivity to the distortion was negative, suggesting that performance improved when intensity perturbation was applied. However, the negative sensitivity may be explained by order and list effects. As indicated before, to allow comparison between listeners, subjects listened to the same lists in each condition, in the same order. As a result, order and list effects may be present in the data across tests.

Even though the individual standard errors are large, a trend can be observed in Fig. 4.6: a decrease in sensitivity as the SII-SR*B*T increases. A linear regression analysis on the data of the hearing-impaired listeners showed a significant correlation of -0.54 ($p<0.05$). From this it may be concluded that the higher the SII (more severe speech processing deficits) the less sensitive the hearing-impaired listeners are to the intensity distortion.

In summary, the results provide evidence that speech intelligibility for the group of hearing-impaired listeners with suprathreshold deficits is affected less by intensity perturbation than for normal-hearing listeners. Moreover, looking at the individual results of all hearing-impaired listeners, the sensitivity to the intensity perturbation correlates negatively with the SII-SR*B*T. In other words, the larger the effect of suprathreshold deficits on speech processing, the less sensitive a hearing-impaired listener is to intensity perturbation. Under the distortion-sensitivity model, this implies that distortion of intensity coding relates to the effects of suprathreshold deficits underlying the poor

speech intelligibility in noise. The underlying deficit might be poor auditory intensity coding.

III. SUMMARY AND CONCLUSIONS

In this study, speech intelligibility was measured as a function of intensity perturbation of speech-plus-noise stimuli. The sensitivity to the distortion by hearing-impaired listeners was compared with that by normal-hearing listeners. The data on the speech intelligibility tests were converted to SII-values. An elevation of the SII of a hearing-impaired listener, as compared with the SII's of normal-hearing listeners, indicates a suprathreshold speech processing deficit; the higher the SII, the more speech intelligibility is affected by suprathreshold deficits. The hearing-impaired listeners were divided into two groups on the basis of their SII-SR*B*T: a group *with* and a group *without* suprathreshold deficits. This classification did not relate to hearing loss: some listeners with a severe hearing loss did not show suprathreshold deficits, whereas some listeners with a mild hearing loss showed severe suprathreshold deficits. Data revealed that hearing-impaired listeners *without* suprathreshold deficits were just as sensitive to intensity perturbations as normal-hearing listeners; hearing-impaired listeners *with* suprathreshold deficits appeared to be less sensitive to intensity perturbations than normal-hearing listeners. The convergence for increasing degrees of intensity perturbation suggests that the applied artificial distortion relates to the suprathreshold deficit causing speech intelligibility problems. A small but significant correlation between the SII-SR*B*T of hearing-impaired listeners and the sensitivity to the intensity distortion was observed. It is concluded that intensity perturbation may partly characterize the effect of a suprathreshold deficit causing a reduced speech intelligibility in noise. The underlying hearing deficit may be a reduced acuity of auditory intensity coding.

REFERENCES

Allen, J. B. (**1977**). "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," IEEE Trans. Acoust. Speech Signal Process. **25**, 235–238.

Allen, J. B., and Rabiner, L. R. (**1977**). "A unified approach to short-time Fourier analysis and synthesis," Proc. of the IEEE **65**, 1558–1564.

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

Buus, S., Florentine, M., and Zwicker, T. (**1995**). "Psychometric functions for level discrimination in cochlearly impaired and normal listeners with equivalent-threshold masking," J. Acoust. Soc. Am. **98**, 853–861.

Florentine, M., Reed, C. M., Rabinowitz, W. M., Braida, L. D., and Durlach, N. I. (**1993**). "Intensity perception. XIV. Intensity discrimination in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **94**, 2575–2586.

Gabor, D. (**1947**). "Acoustical quanta and the theory of hearing," Nature (London) **159**, 591–594.

Houtgast, T. (**1995**). "Psycho-acoustics and speech recognition of the hearing impaired," in *Proceedings of the European Conference on Audiology*, Noordwijkerhout, The Netherlands, pp. 165–169.

Moore, B. C. J. (**1995**). *Perceptual consequences of cochlear damage*, (University Press, Oxford).

Moore, B. C. J. (**1996**), "Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids," Ear Hear. **17**, 133–161.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**1999**). "Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. I. Normal-hearing listeners," J. Acoust. Soc. Am. **105**, 2895–2902.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**in press**). "Measuring the threshold for speech-reception by adaptive variation of the signal bandwidth. II. Hearing-impaired listeners," to appear in J. Acoust. Soc. Am. .

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the Speech Reception Threshold for sentences," Audiology **18**, 43–52.

Rioul, O., and Vetterli, M. (**1991**). "Wavelets and signal processing," IEEE Signal Proc. Mag. October, 14–38.

Scharf, B. (**1970**). "Critical bands," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. 1, pp. 157–202.

Smoorenburg, G. F. (**1992**). "Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram," J. Acoust. Soc. Am. **91**, 421–437.

van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**1999**). "Intensity discrimination of Gaussian-windowed tones: Indications for the shape of the auditory frequency-time window," J. Acoust. Soc. Am. **105**, 3425–3435.

5 Effects of degradation of intensity, time, or frequency content on speech intelligibility for normal-hearing and hearing-impaired listeners

Many hearing-impaired listeners suffer from distorted auditory processing capabilities. This study examines which aspects of auditory coding (i.e., intensity, time, or frequency) are distorted and how this affects speech perception. The distortion-sensitivity model is used: the effect of distorted auditory coding of a speech signal is simulated by an artificial distortion, and the sensitivity of speech intelligibility to this artificial distortion is compared for normal-hearing and hearing-impaired listeners. Stimuli (speech plus noise) are wavelet coded using a complex sinusoidal carrier with a Gaussian envelope (¼ octave bandwidth). Intensity information is distorted by multiplying the modulus of each wavelet coefficient by a random factor. Temporal and spectral information are distorted by randomly shifting the wavelet positions along the temporal or spectral axis, respectively. Measured were (1) detection thresholds for each type of distortion, and (2) speech-reception thresholds (SRTs) for various degrees of distortion. For spectral distortion, hearing-impaired listeners showed increased detection thresholds and were also less sensitive to the distortion with respect to speech perception. For intensity and temporal distortion, thresholds and sensitivity both were normal. Results indicate that a distorted coding of spectral information may be an important factor underlying reduced speech intelligibility for the hearing impaired.

Submitted to the Journal of the Acoustical Society of America

INTRODUCTION

The difficulty hearing-impaired listeners have to perceive speech in noise has been the subject of many investigations, but is still not entirely understood. Although audibility plays an important role, several studies have shown that this cannot explain the whole problem [see, for example, Moore (1996) or Noordhoek *et al.* (in press)]. These studies have demonstrated that factors apart from reduced audibility, called suprathreshold deficits, degrade speech processing. Suprathreshold deficits can distort the auditory processing of either intensity, time, or frequency information, or a combination of these types of information. For example, excessive forward and backward masking are consequences of suprathreshold deficits that may be reduced a single factor of distorted temporal coding; excessive upward and downward spread of masking may be related to distorted spectral coding. Impaired loudness perception probably relates to a distorted representation of intensity information. This study evaluates these three types of information. The aim is to investigate how reduced speech intelligibility relates to distorted coding of intensity, time, or frequency.

Auditory coding cannot be manipulated directly. However, one can investigate the differences in auditory functions among hearing-impaired subjects on specific auditory tests related to accuracy of intensity, time or frequency coding, and correlate these with their speech perception performance. In several studies this correlation approach was applied, concentrating on the role of reduced temporal or spectral resolution. The role of reduced temporal resolution in reduced speech intelligibility in noise is not yet clear. In some studies a significant correlation between speech intelligibility and temporal resolution was found (Tyler *et al.*, 1982; Dreschler and Plomp, 1985; Moore and Glasberg, 1987); in other studies this was not so (Festen and Plomp, 1983; van Rooij and Plomp, 1990). With respect to reduced spectral resolution, in most studies a significant correlation with speech intelligibility was found (Patterson *et al.*, 1982; Festen and Plomp, 1983; Dreschler and Plomp, 1985; Horst, 1987). On the other hand, this was not the case in a few other studies (van Rooij and Plomp, 1990; Smoorenburg, 1992).

The correlation approach results in statistical relations between reduced speech perception and suprathreshold deficits. A drawback of this approach is that one cannot

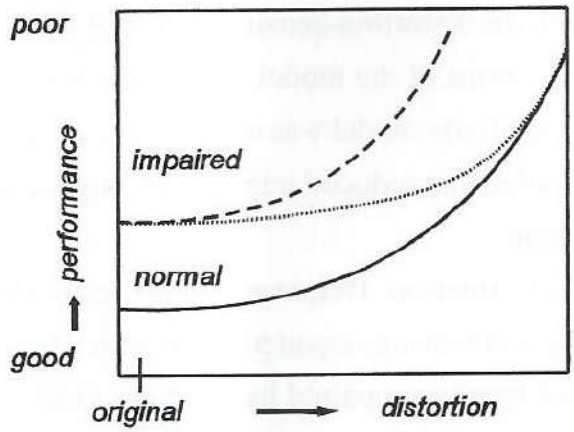


FIG. 5.1. Illustration of the distortion-sensitivity model. Performance for hearing-impaired listeners as a function of distortion is compared with that of normal-hearing listeners (solid line). The possible outcome of such an experiment is "convergence" (dotted and solid lines) or "no convergence" (dashed and solid lines).

exclude that an underlying common factor causes the observed correlation. For example, if a correlation between speech intelligibility and spectral resolution is observed, an underlying common factor can be the hearing threshold. Then, higher hearing thresholds instead of reduced frequency selectivity may cause reduced speech perception. In different studies, underlying factors probably had different effects, which may explain the different results. Relations between distorted auditory coding and speech perception can be investigated in a more direct way using the distortion-sensitivity model (Houtgast, 1995; Chapter 4 of this thesis).

Under the distortion-sensitivity model (Fig. 5.1), the relation between speech intelligibility and a distorted *auditory* coding is studied by simulating the effect of the auditory deficit by *artificial* distortion of the speech signal. The idea is that removing cues that are not perceived by the hearing impaired will not affect their performance. Performance is measured as a function of distortion, and compared for normal-hearing and hearing-impaired listeners. Two trends may be observed: *convergence* (dotted and solid lines) or *no convergence* (dashed and solid lines). In the convergence case, hearing-impaired listeners are less sensitive to the distortion than normal-hearing listeners. Then, it may be concluded that the artificial distortion relates to distorted auditory coding that impedes performance. The artificial distortion affects the sound characteristics in the same way as the auditory deficits. In the no-convergence case, hearing-impaired listeners are as sensitive to the distortion as normal-hearing listeners, indicating that the artificial distortion has no relation to hearing deficits causing difficulties in speech perception. A

few studies (Duquesnoy and Plomp, 1980; ter Keurs *et al.*, 1993; Turner *et al.*, 1995; Chapter 4 of this thesis) used the principles of the distortion-sensitivity model so far, but they did not explicitly explain their results in terms of the model, except the last study.

In Chapter 4 of this thesis the distortion-sensitivity model was used with respect to the coding of intensity information. It was concluded that reduced intensity coding accuracy may partly explain impaired speech perception.

With respect to the coding of temporal information, Duquesnoy and Plomp (1980) measured speech reception of normal-hearing and hearing-impaired listeners as a function of reverberation time. Their results show that hearing-impaired listeners are as sensitive to reverberation as normal-hearing listeners. In terms of the distortion-sensitivity model, this leads to the conclusion that speech perception problems are not caused by a deficit that introduces a delay to parts of the speech energy, as distorted temporal coding may do.

With respect to coding of spectral information, ter Keurs *et al.* (1993) compared the effect of reduced spectral contrast on speech perception in normal-hearing and hearing-impaired listeners. They concluded that "limited resolution of spectral contrast is only loosely associated with hearing loss for speech in noise." Turner *et al.* (1995) compared speech reception of hearing-impaired and normal-hearing listeners for unprocessed speech and for speech in which spectral cues were removed. For the original speech, hearing-impaired listeners had lower speech-intelligibility scores than the normal-hearing listeners. However, for speech without spectral cues, hearing-impaired listeners understood as well as normal-hearing listeners. In terms of the distortion-sensitivity model, this convergence indicates that the reduced speech intelligibility by hearing-impaired listeners is related to a degraded processing of spectral cues. It should be mentioned that this is our interpretation of the data. Turner *et al.* were interested in the ability of hearing-impaired listeners to use temporal cues. Their conclusion, not in conflict with ours, is that the temporal accuracy of speech coding of hearing-impaired listeners is not impaired in terms of speech recognition.

The studies mentioned above obtained data that can be analyzed in terms of the distortion-sensitivity model. The effects of distortion of intensity, time, and frequency information on speech perception were studied in isolation, although these three domains are not completely independent. Manipulation in one domain will affect the other domains. For example, spectral smearing introduces temporal smearing and vice versa.

In Sec. 5.I A, this will be illustrated. Being aware of these unwanted byproducts of the speech processing algorithm is important. Therefore, in the present study, the interdependency of the intensity, time, and frequency domains was taken into account.

In short, this study addresses which domains in auditory coding (i.e., intensity, time, or frequency) cause speech-perception problems for hearing-impaired listeners. First, it is investigated which sound domains are less clearly perceived by hearing-impaired listeners. For this, detection thresholds for artificially applied distortions of intensity, time, or frequency are measured. If a particular type of information is less clearly perceived by hearing-impaired listeners, the detection thresholds for the distortion of this information will probably be higher. The influence of distorted coding on speech perception was investigated by means of the distortion-sensitivity model. Speech intelligibility is measured as a function of the degree of artificial distortion of intensity, time, or frequency information. Comparison of the performance for normal-hearing and hearing-impaired listeners may provide insight into the role of reduced accuracy in auditory coding as a possible explanation for the degraded performance of the hearing impaired.

I. METHOD

A. Degradation of intensity, time, and frequency information

In this study, a sound processing algorithm is used to degrade artificially the intensity, time, and frequency content of speech. The degradation is intended to simulate the effects of distorted auditory coding. By means of the speech-reception threshold test (SRT, Sec. 5.I D3), speech intelligibility of sentences was measured as a function of applied artificial distortion. In order to simulate auditory coding, a perceptually relevant spectro-temporal decomposition and recomposition method was developed. This method was also used in Chapter 4, and is described below.

- **Spectro-temporal decomposition & recomposition**

To model auditory spectro-temporal coding, sounds were described in the time-frequency domain by means of a wavelet transform. Compared with the short-time Fourier transform, the wavelet transform matches auditory system coding more closely because it uses a logarithmic frequency scale (e.g., Rioul and Vetterli, 1991). An important criterion in the choice of the mother wavelet is its spectral and temporal width. Results of Chapters 2 and 3 suggest that a Gaussian-windowed sinusoid with a shape factor between 0.15 and 0.3 roughly matches the auditory time-frequency window. Therefore, as the prototype analysis function, a Gaussian wavelet was chosen. The Gaussian wavelet is a complex sinusoidal carrier with a Gaussian envelope:

$$s(t) = \sqrt{\alpha f_0}\, \exp(i2\pi f_0 t)\, \exp(-\pi(\alpha f_0 t)^2) \quad , \tag{5.1}$$

in which f_0 is the carrier frequency, α is the shape factor, and $\sqrt{\alpha f_0}$ normalizes the energy of the analysis function. This time-frequency window has an effective bandwidth of $\Delta_f=\alpha f_0$ and an effective duration of $\Delta_t=1/(\alpha f_0)$ (Gabor, 1947). The effective bandwidth of the analysis function was set to ¼ octave [roughly equal to the auditory critical band (Scharf, 1970)]. This corresponds with a shape factor $\alpha=0.1735$. As a result, the effective duration of the time-frequency window is 5.76 ms at 1 kHz (1.44 ms at 4 kHz). The effective number of periods contained within the Gaussian envelope equals 5.8 (i.e., $1/\alpha$).

This Gaussian wavelet was used to construct a wavelet decomposition that covers the time-frequency plane. Shifts of this prototype analysis function cover the temporal range; scales of the prototype function cover the spectral range. The scaling is controlled by varying the carrier frequency f_0. The decomposition results in complex wavelet coefficients, which can be characterized by a modulus, a phase, and a position in the time-frequency plane.

For simultaneous sampling in time and frequency the Nyquist sampling theorem was applied twice (Allen, 1977; Allen and Rabiner, 1977). The sampling interval was based on the temporal and spectral range over which the Gaussian wavelet is essentially different from zero. Since the Gaussian wavelet does not have compact support[1] in time,

---

[1]A function $f(t)$ has compact support if it is zero outside the interval $T_0<t<T_0+\Delta T$.

nor in frequency, the range between the points that were 25 dB down from the peak was taken as the range over which the window is significant (about twice the effective duration and effective bandwidth). This criterion leads to a sampling of one wavelet per three periods of the wavelet carrier frequency along the time axis, and eight wavelets per octave along the frequency axis. Theoretically, the number of complex coefficients needed to describe the signal using the 25-dB criterion for sampling, is about two coefficients per input sample (Allen, 1977). In this study, the frequency of the signals was limited to the range from 250 to 4000 Hz. As a result, one second of speech (sampling frequency: 44.1 kHz; no information below 250 Hz or above 4 kHz preserved) was described by $16*10^3$ complex wavelet coefficients.

Using these wavelet coefficients, sounds can be reconstructed by an overlap-add procedure. Theoretically, the reconstruction is not perfect. However, using the 25-dB criterion for sampling in time and frequency, little or no aliasing occurs in either the time or the frequency domain. Adequate sampling is important for two reasons (Allen and Rabiner, 1977). First, the difference between the recomposed signal and the original signal must not be noticeable to a listener. Second, in this study modifications to the spectro-temporal decomposition of sound are performed. When modifying undersampled spectro-temporal representations of sound, interactions between modification and window shape may occur. Such interactions will lead to unwanted byproducts. As a result of the careful sampling in our decomposition and recomposition scheme, (1) the difference between an original and a recomposed signal was very small and not noticeable to the listener, and (2) the scheme is robust for interactions between window shape and modifications of the decomposition.

Between decomposition and recomposition, the accuracy of the intensity, time, or frequency information was degraded to simulate poor auditory coding. Intensity degradation was obtained by introducing uncertainty in the modulus of each wavelet coefficient. Temporal and spectral degradations were obtained by introducing uncertainty in the temporal and spectral position of each wavelet, respectively. In Fig. 5.2, this is illustrated schematically. In the following paragraphs, these different types of degradation will be explained in more detail. After the perturbation, the energy contained in each frequency band over the whole test sentence was scaled to equal the original energy in that band. Since this study aims at investigating speech perception performance in noise, speech and noise were summed before processing.

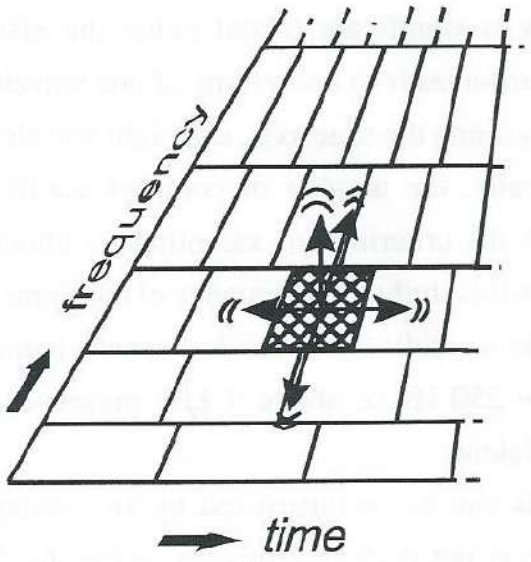


FIG. 5.2. Schematic illustration of the perturbation of the intensity, time, or spectral information. The Gaussian wavelets are symbolized by rectangles. Each wavelet is given a random perturbation with respect to its intensity, temporal position, or spectral position.

- **Degradation of the intensity accuracy**

To degrade the accuracy of the intensity information, the modulus of the wavelet coefficients was perturbed (intensity perturbation). This was achieved by multiplying each wavelet coefficient by a random factor. As a result, silence will remain silence after perturbation. The random perturbation factor ε (in dB) was chosen from a uniform distribution with zero mean and boundaries[2] $-I_D/2$ and $+I_D/2$. Thus the modulus of each individual coefficient was multiplied by a different random factor $10^{\varepsilon/20}$.

- **Degradation of the temporal accuracy**

To degrade the accuracy of the temporal information, the positions of the wavelets were shifted randomly along the temporal axis (temporal perturbation). To avoid a degradation of the accuracy of spectral information as much as possible, only the temporal envelope of the wavelets was displaced, not the underlying fine structure. The new fine structure was calculated by extrapolation of the original fine structure to the new position of the envelope. As a result, the information contained within the original fine structure was left unaffected. The position of the envelope of each wavelet was shifted independently by a random value chosen from a uniform distribution ranging from $-T_D/2$ to $+T_D/2$. The

[2]In Chapter 4, the random perturbation factor with which the modulus of each wavelet coefficient was multiplied was chosen from a uniform distribution with boundaries $-I_D$ and $+I_D$.

degree of temporal distortion T_D is expressed in terms of the duration of the wavelets (inversely proportional to the bandwidth). If T_D equals two wavelets, the maximal displacement along the time axis is one effective duration of the wavelet from its original position. At 1 kHz, this is 5.76 ms; at 4 kHz, this is 1.55 ms.

- **Degradation of the spectral accuracy**

To degrade the accuracy of the spectral information, the position of each wavelet was shifted randomly along the spectral axis (spectral perturbation). The positions of all wavelet coefficients were shifted independently by a random value chosen from a uniform distribution ranging from $-F_D/2$ to $+F_D/2$. The degree of spectral distortion F_D is expressed in octaves. If $F_D = 0.5$ octaves, the maximal displacement along the frequency axis is 0.25 octaves (equals the effective bandwidth of the analysis window).

After wavelet decomposition, the spectral information is not only encoded in the position of the wavelets along the spectral axis, but also in the phase of the coefficients. The relative phases of the coefficients in each frequency band contain information about the spectral structure within this band. The random shifts of the wavelet positions along the spectral axis result in a smeared spectrum over bands. However, if the phase is kept intact, part of the spectral information within a band is reintroduced in the overlap-add procedure by interactions between neighboring wavelets.[3] By distorting the phase information we tried to bypass this problem. The phase was distorted by a desynchronization of the regular pattern of the wavelet coefficients along the temporal axis. This desynchronization was obtained by shifting the position of each wavelet (envelope plus fine structure) along the temporal axis by a random value chosen from a uniform distribution ranging from -0.0375 to +0.0375 of the wavelet bandwidth. In all conditions with spectral distortion including the spectral reference condition (0-octaves spectral perturbation), the phase was distorted in this way.

In Fig. 5.3, the effect of distorting the spectral information of an artificial vowel /a/ is illustrated. Panel a shows the undistorted vowel. In panel b, the vowel is plotted in the spectral reference condition. In this condition, the phase of the complex coefficients is

---

[3]This inherent characteristic of overlap-add procedures was described in more detail by Baer and Moore (1993). Without phase distortion, even for large random shifts along the spectral axis, basic periodicity in the spectrum is preserved due to the preserved coherence of the phase spectrum.

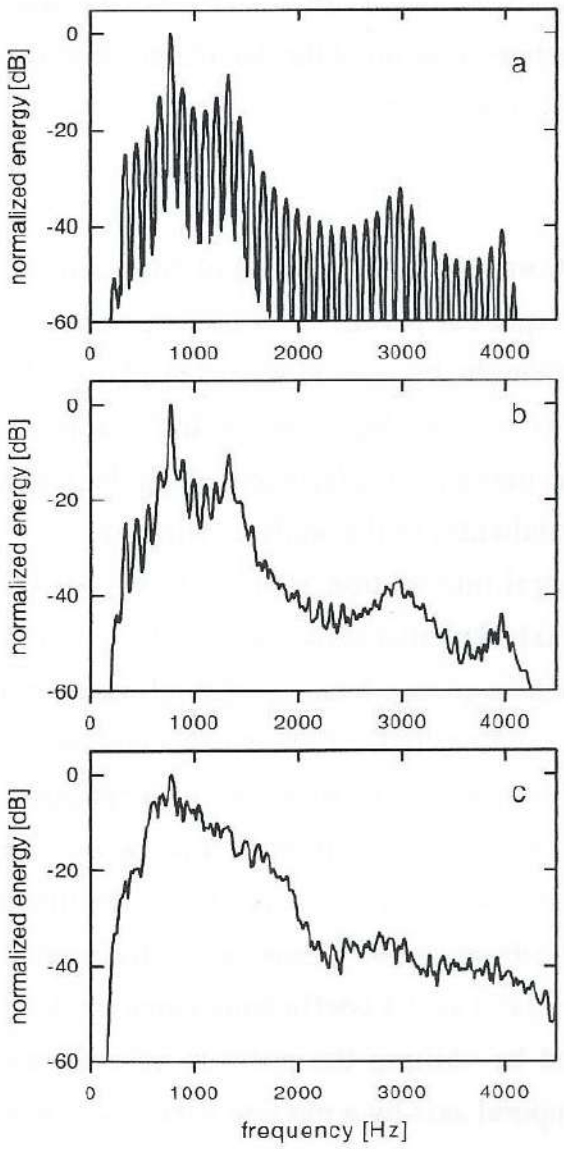


FIG. 5.3. The effect of the artificial distortion of the spectral information on an artificial vowel /a/. (a) undistorted vowel; (b) spectral reference condition (phase distorted) (c) spectral perturbation of 0.75 octaves (phase distorted and spectrally perturbed).

distorted, but the positions of the wavelets along the spectral axis are retained. As a result, most of the spectral fine structure is lost, but the spectral envelope is intact. In panel c, the vowel is plotted in the most severe spectral distortion condition used in this study, i.e., when F_D equals 0.75 octaves. The phase is distorted as in the reference condition, and in addition the wavelets were shifted randomly over maximal $F_D/2$ along the spectral axis. As a result, the spectral envelope is smeared almost fully. Thus the overall spectral effect of the applied spectral uncertainty is a broadening of the spectral peaks.

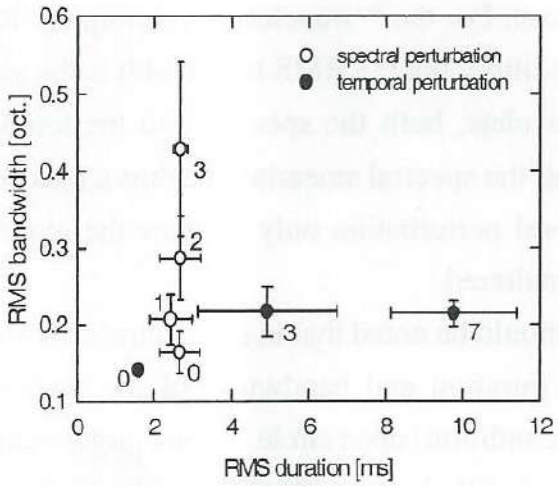


FIG. 5.4. The effect of the non-deterministic perturbation process on the RMS duration[4] and RMS bandwidth of a Gaussian-windowed tone with a center frequency of 1 kHz and a shape factor of 0.1735, i.e., an effective bandwidth of ¼ octave. Filled and open symbols represent the values corresponding with temporal perturbation and spectral perturbation, respectively. The numbers represent the degree of perturbation (expressed in the number of wavelets). The error bars represent the standard deviation.

As mentioned in the Introduction, degradation of the accuracy of the information of one domain is not possible without collateral degradation of the information of other domains. For example, the degradation of the accuracy of the intensity information also affects the spectral and temporal content of a signal. The effects of distortion of temporal information on spectral information and vice versa are illustrated in Fig. 5.4 for a Gaussian-windowed sinusoid as input to the wavelet decomposition, followed by spectral or temporal degradation, and recomposition. The RMS duration[4] and RMS bandwidth of this Gaussian-windowed sinusoid (center frequency = 1 kHz; $\alpha = 0.1735$) are indicated by the filled circle with index '0'. The effects of temporal perturbation on the duration and bandwidth of the signal are represented by the other filled circles; the effects of the spectral perturbation are indicated by open circles. The perturbation procedure was applied to the input signal six times. The error bars represent the standard deviations of the resulting duration and bandwidth of the output signals.

---

[4]The root mean square (RMS) duration of a function $f(t)$ is defined by

$$\Delta_t := \frac{1}{\|f(t)\|_2}\sqrt{\int_{-\infty}^{\infty} t^2 f^2(t)\,dt}\ .$$

The RMS bandwidth is defined analogously.

Looking at the effect of temporal perturbation, it can be observed that, when a temporal perturbation of 3 wavelets is applied, both the RMS duration and RMS bandwidth of the Gaussian tone pulse increase. For the 7-wavelets condition, the RMS duration is longer than in the 3-wavelets condition, but the RMS bandwidth is the same. Thus for temporal perturbation up to 3 wavelets, both the spectral and the temporal contrasts of sound are reduced. At that point, the spectral smearing reaches a maximum of about 0.25 octaves. Beyond that, temporal perturbation only reduces the temporal contrasts while the spectral contrasts stay unaltered.

With respect to spectral perturbation, it should be noted that in all spectral conditions the phase was distorted. As a result, the duration and bandwidth of the Gaussian-windowed sinusoid in the spectral reference condition (open circle '0') are larger than the duration and bandwidth of the original signal (filled circle '0'); the spectral reference condition is slightly spectro-temporally smeared. The effect of additional spectral perturbation is just a reduction of the spectral contrasts, while the resulting (after phase distortion) temporal contrasts are maintained.

B. Subjects

Twelve normal-hearing listeners, aged 20 to 63 years with a mean age of 26 years, participated in the experiment. Pure-tone air-conduction thresholds of the normal-hearing listeners did not exceed 15 dB HL at any octave frequency from 250 to 4000 Hz. In addition, twenty-six sensorineurally hearing-impaired listeners took part in the experiment, aged 24 to 67 years with a mean age of 48 years. Their intelligibility scores for monosyllabic words in quiet were at least 75% correct. The pure-tone, air-conduction threshold in the hearing-impaired listener's better-hearing ear was at least 30 dB HL at one or more frequencies between 250 and 4000 Hz. Thresholds of the better-hearing ear averaged over 0.5, 1, and 2 kHz (the pure-tone average, or PTA) ranged from 17 to 70 dB HL, with a mean PTA of 50 dB HL. All listeners were native Dutch speakers.

C. Stimuli and apparatus

The speech stimuli consisted of sentences and words. The sentence sets contained lists of 13 everyday Dutch sentences of eight to nine syllables read by a female and male

speaker (Versfeld *et al.*, in press). The word sets consisted of lists of balanced meaningful CVC-words (Bosman and Smoorenburg, 1995).

Signals were played out over TDT (Tucker Davis Technologies) System II hardware. Stimuli were presented in the middle of the dynamic range of each listener by frequency shaping them using a programmable filter (TDT PF1). The stimuli were presented monaurally through Sony MDR-V900 headphones. To avoid the risk of cross-hearing, the listener's better-hearing ear was tested. For calibration, sound pressure levels of the stimuli were measured on a Brüel & Kjær type 4152 artificial ear with a flat-plate adapter. The entire experiment was controlled by a personal computer. Subjects were tested individually in a soundproof room.

D. Procedures

First, the hearing threshold and the uncomfortable loudness level (UCL) of each listener were determined. In the detection and intelligibility tests, sounds were adapted to fit the dynamic range of each listener. To familiarize the subjects with the procedure, a training session preceded data collection. All conditions were measured twice in order to determine measurement reliability. Speech intelligibility tests were performed once using sentences spoken by the female talker and once using those by the male talker. In the distortion-sensitivity model, the performance for individual hearing-impaired listeners is compared with that for normal-hearing listeners. Therefore, for all listeners, the same order of conditions and sentence lists was used.

- **Threshold and UCL**

The dynamic range of each listener was estimated by measuring the hearing threshold and the uncomfortable loudness level (UCL) for narrow bands of noise. The UCL was corrected for broadband stimulation, as described below.

Thresholds and UCLs were measured using 1/3-octave noise bands at center frequencies of 250, 500, 1000, 2000, and 4000 Hz. Hearing thresholds were measured using a Békésy tracking (Yantis, 1994) procedure (300-ms noise bursts; repetition rate 2.5 Hz; step size 1 dB). The measurement was ended after eleven level reversals. The average of all but the first reversal level was taken as the hearing threshold. Narrow-band UCLs were measured with 1/3-octave noise bursts that were presented with a 3-dB

increase in level for each presentation (300 ms noise burst; repetition rate 1.4 Hz). Listeners were asked to press a button when the noise bursts became uncomfortably loud. Then, the level of the noise burst was immediately diminished by a random amount between 21 and 30 dB, and the ascending procedure was repeated until six responses were obtained. The average of the levels at which the button was pushed was taken as the narrow-band UCL.

To correct the UCL for broadband stimulation, a 4-second broadband noise burst was presented, spectrally shaped according to the narrow-band UCLs and starting 40 dB below the narrow-band UCLs. The level of the broadband noise burst was gradually increased in steps of 5 dB. After each presentation the listener was asked whether the signal was experienced as uncomfortably loud. If this was the case, the corresponding level was taken as the broadband UCL.

- **Detection threshold for distortion**

The detection thresholds for the distortion of intensity, temporal, or spectral information were estimated using words. A 3I-3AFC two-down one-up adaptive procedure was used, leading to a 70.7 % correct score. In each trial, the subject was presented with three signals, twice the reference word and once the distorted word. The listener had to point out the distorted one. For each trial, a random choice out of 90 bandpass filtered (250-4000 Hz) pre-processed (at different degrees of distortion) words was loaded from disk. The difficulty of the task was increased by dividing the distortion factor by $\sqrt{2}$ following two consecutive correct responses; the difficulty of the task was decreased by multiplying the distortion factor by $\sqrt{2}$ following one incorrect response. A transition from increasing to decreasing difficulty or vice versa defined a reversal. A run was ended after 20 reversals. The geometric mean of the last 16 reversals was used as an estimate of the detection threshold for distortion. To define the experiment with respect to presentation level, all words were presented in the middle of the dynamic range of the listener, in speech noise (Wandel und Goltermann RG-1) at a signal-to-noise ratio of 15 dB.

- **Speech intelligibility**

Speech-reception threshold in noise for an adapted spectrum (SRTa)

The speech-reception threshold (SRT) is an estimate of the ability to perceive speech in daily life (Plomp and Mimpen, 1979). The SRT in noise is defined as the signal-to-noise ratio (SNR) at which 50% of the sentences are reproduced correctly. The speech level is varied in an adaptive, up-down procedure with a step size of 2 dB. The continuous stationary noise is presented from 500 ms before to 500 ms after the sentence. In our experiments, speech and noise are adapted to fit in the dynamic range of individual listeners. This adapted speech-reception threshold is called SRTa. In the SRTa tests in this study, all stimuli were bandpass filtered from 250 to 4000 Hz.

After an SRT test using undistorted speech, the SRTa was measured as a function of the degree of distortion (distortion-sensitivity model). The intensity-distortion conditions were 0 (undistorted), 10, and 20 dB. The temporal-distortion conditions were 0 (undistorted), 3, and 7 wavelets. The spectral-distortion conditions were 0, ¼, ½, and ¾ octave (recall that in all spectral-distortion conditions the phase was distorted).

Speech-Reception Bandwidth Threshold (SRBT)

In addition to the SRTa, the Speech-Reception *Bandwidth* Threshold (SR*B*T) was measured to estimate suprathreshold speech processing. The SR*B*T measure of speech intelligibility was introduced by Noordhoek *et al.* (1999). The SR*B*T is highly sensitive for suprathreshold deficits, as is shown in a recent study of Noordhoek *et al.* (in press).

The SR*B*T procedure is similar to the SRT procedure, except that the bandwidth (center frequency: 1 kHz) of the undisturbed speech is varied instead of the level when estimating the 50% intelligibility threshold. Complementary shaped bandstop noise is added to the bandpass-filtered speech. Speech and noise are presented in the middle of the listener's dynamic range.

E. Speech Intelligibility Index

To estimate the quality of speech processing of listeners, the SRTa and SR*B*T data were converted to a Speech Intelligibility Index. The Speech Intelligibility Index (SII) (ANSI, 1997) is a physical measure of how much information of speech is available to the

listener. The SII correlates highly with speech intelligibility. To perceive speech, normal-hearing listeners need a certain amount of information which can be converted to an SII value. If hearing-impaired listeners need more information, this suggests that their speech processing is degraded. Thus elevated SII values are an indication for a low speech processing quality. The SII model accounts for hearing threshold, self-masking in speech, normal upward spread of masking and level distortion at high presentation levels. To calculate the SII, speech spectra, noise spectra, and hearing thresholds must be known. As mentioned in Sec. 5.I D1, hearing thresholds were measured with 1/3-octave noise bands, using Békésy tracking (Yantis, 1994). This procedure probably results in hearing thresholds that are systematically about 4 dB higher than the methods on which the ISO (1961) threshold is based (Noordhoek *et al.*, in press; Noordhoek *et al.*, submitted). Therefore, in the SII calculations the internal noise level was lowered by 4 dB. The band-importance function for speech material of average redundancy (Pavlovic, 1987) was used.

II. RESULTS AND DISCUSSION

A. Detection thresholds

To obtain insight into which attributes of sound processing are distorted for hearing-impaired listeners, detection thresholds for the distortion of intensity, time, and frequency information were measured. If the auditory coding of a particular type of information is degraded, the detection thresholds for the distortion of this type of information are assumed to be higher.

- **Degradation of the intensity accuracy**

For the normal-hearing listeners, the detection threshold for the intensity perturbation, described in Sec. 5.I A, ranged from 13 to 23 dB, with a median of 17 dB. For the hearing-impaired listeners, the detection thresholds ranged from 9 to 53 dB, with a median of 18 dB. The overall (normal-hearing plus hearing-impaired listeners: 38

subjects) mean standard error of an individual detection threshold (2 measurements) was 3 dB. A Mann-Whitney *U* test showed that the difference in detection threshold between normal-hearing and hearing-impaired listeners was not significant.

- **Degradation of the temporal accuracy**

For the normal-hearing listeners, the detection thresholds for temporal perturbation ranged from 0.9 to 1.5 wavelets, with a median of 1.1 wavelets; for the hearing-impaired listeners, this threshold ranged from 0.6 to 7.4 wavelets, again with a median of 1.1 wavelets. The mean standard error of an individual detection threshold was 0.4 wavelets. A Mann-Whitney *U* test showed that the detection thresholds for the group of hearing-impaired listeners were not significantly higher than those for the normal-hearing listeners.

- **Degradation of the spectral accuracy**

For the normal-hearing listeners, the detection thresholds for spectral perturbation ranged from 0.22 to 0.39 octave, with a median of 0.26 octave. For the hearing-impaired listeners, the detection thresholds ranged from 0.17 to 1.4 octave, with a median of 0.36 octave. The mean standard error of the individual detection threshold was 0.06 octave. A Mann-Whitney *U* test showed that the detection thresholds for the group of the hearing-impaired listeners were significantly ($p<0.05$) higher than those for the normal-hearing listeners.

In summary, with respect to the detection of distortion of intensity and temporal information, no significant difference was observed between the group of normal-hearing and the group of hearing-impaired listeners. With respect to the detection of spectral distortion, a significant difference between normal-hearing and hearing-impaired listeners was observed. Thus spectral cues were probably less clearly perceived by the hearing-impaired listeners.

B. Suprathreshold Speech Intelligibility

The aim of this study is to gain insight into the suprathreshold speech processing problems of hearing-impaired listeners. Therefore, speech processing performance was

measured by means of the SRTa and SR*B*T test. For the normal-hearing listeners, the SRTa ranged from -1.8 to 0.3 dB, with a median of -0.8 dB. For the hearing-impaired listeners, the SRTa ranged from -1.1 dB to 8.5 dB, with a median of 2.0 dB. The mean standard error of an individual SRTa (six measurements) was 0.7 dB. The hearing-impaired listeners had significantly higher SRTa's than the normal-hearing listeners (Mann-Whitney *U* test: $p<0.05$). The SR*B*T for the normal-hearing listeners ranged from 1.1 to 1.7 octave, with a median of 1.6 octave. The SR*B*T for the hearing-impaired listeners ranged from 1.5 to 3.4 octave, with a median of 2.1 octave. The standard error of an individual SR*B*T (2 measurements) was 0.3 octave. The hearing-impaired listeners had significantly higher SR*B*T values than the normal-hearing listeners (Mann-Whitney *U* test: $p<0.05$).

For both the SRTa and the SR*B*T tests, hearing-impaired listeners performed worse than normal-hearing listeners, which confirms the problems hearing-impaired listeners have in perceiving speech. To quantify the degree of deterioration of suprathreshold speech processing, the individual SRTa and SR*B*T data were converted to SII units. For the normal-hearing listeners, the SII for the SRTa ranged from 0.36 to 0.42, with a median of 0.39; the SII for the SR*B*T ranged from 0.26 to 0.39, with a median of 0.35. For the hearing-impaired listeners, the SII for the SRTa ranged from 0.37 to 0.54, with a median of 0.43; the SII for the SR*B*T ranged from 0.32 to 0.52, with a median of 0.43. The individual standard error of the SII_{SRTa} (6 measurements) was 0.02. The individual standard error of the SII_{SRBT} (2 measurements) was 0.05. Both the SII_{SRTa} and the SII_{SRBT} for the hearing-impaired listeners were significantly higher than those for the normal-hearing listeners (Mann-Whitney *U* test; $p<0.05$).

The SII values of the hearing-impaired listeners indicate that their suprathreshold speech processing is clearly distorted. The next step is to explore what aspects of auditory coding are distorted. The detection threshold experiments suggest that hearing-impaired listeners perceive spectral information less clearly than normal-hearing listeners. In Fig. 5.5 the individual detection thresholds for spectral perturbation are plotted as a function of the SII_{SRTa} (panel a) and as a function of the SII_{SRBT} (panel b). Open symbols represent the detection thresholds for the normal-hearing listeners, filled symbols those for the hearing-impaired listeners. The figure shows a correlation between the SII's and the detection threshold for spectral perturbation. A statistical analysis (Spearman rank correlation) on the data for the normal-hearing and hearing-impaired listeners confirmed

this: there is a significant ($p<0.005$) correlation of 0.5 between the detection threshold for spectral perturbation and SII_{SRTa}, and a significant ($p<0.0005$) correlation of 0.6 between the detection threshold and SII_{SRBT}.

Summarizing, a correlation between the detection threshold for the distortion of spectral information and the SII was observed. Thus less clearly perceived spectral information by hearing-impaired listeners relates statistically to their speech processing deficits. In the next section, the relation between the auditory coding accuracy and reduced speech intelligibility is analyzed in a more direct way by means of the distortion-sensitivity model.

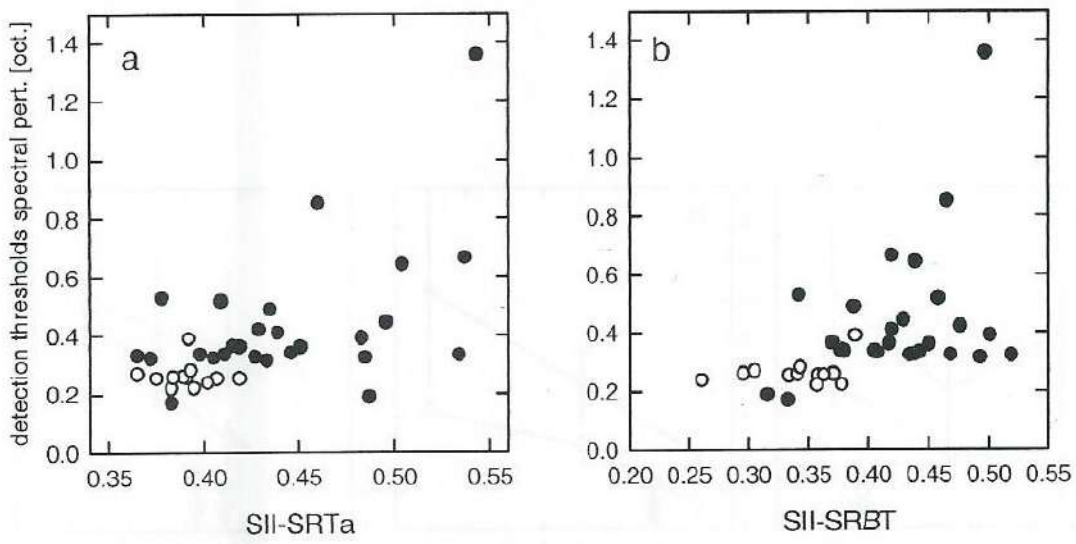


FIG. 5.5. Individual detection thresholds for spectral perturbation versus the Speech Intelligibility Index (SII) corresponding with the mean of SRTa-scores (panel a) and SR*B*T-scores (panel b) for normal-hearing listeners (open circles) and hearing-impaired listeners (filled circles).

C. Distortion-sensitivity model: group results

Applying the distortion-sensitivity model, the SRTa was measured as a function of the artificial degradation of the spectro-temporal coding of sound, for normal-hearing and hearing-impaired listeners. The results are plotted in Fig. 5.6. The SRTa is plotted as a function of the degree of distortion of intensity information (panel a), temporal information (panel b), and spectral information (panel c). Open and filled circles represent the medians of the data for the normal-hearing and hearing-impaired listeners, respectively. The bars represent the inter-quartile ranges. The arrows represent the medians of the detection thresholds for normal-hearing (open circle) and hearing-impaired listeners (filled circle).

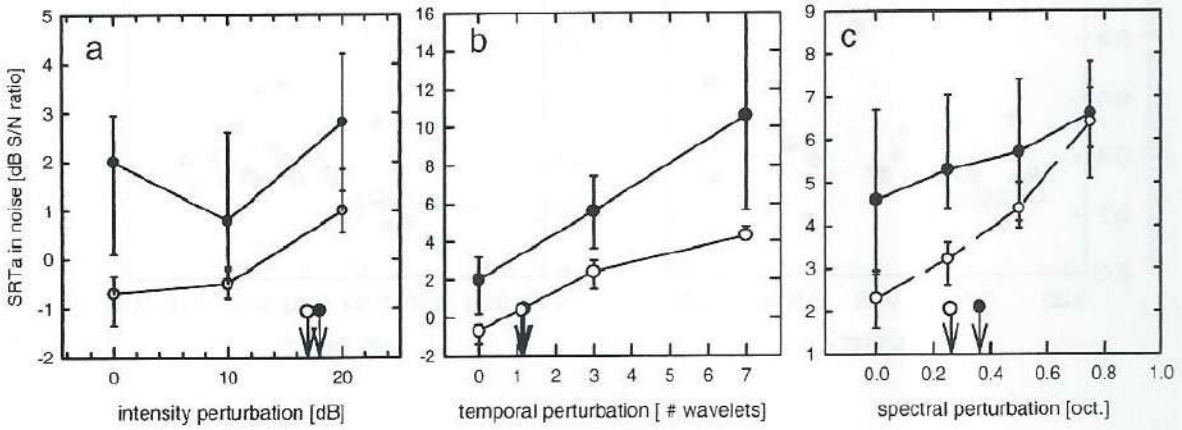


FIG. 5.6. The median of SRTa-values for normal-hearing (open symbols) and hearing-impaired listeners (filled symbols) as a function of distortion. The error bars represent the inter-quartile ranges. Arrows indicate the median of the detection threshold for each distortion for the normal-hearing listeners (open circle) and hearing-impaired listeners (filled circle). Panel a: distortion of intensity information; panel b: distortion of temporal information; panel c: distortion of spectral information.

- **Degradation of the intensity accuracy**

For all levels of intensity degradation, the hearing-impaired listeners perform poorer than the normal-hearing listeners on the speech intelligibility tests (Fig. 5.6a). The difference in performance between normal-hearing and hearing-impaired listeners appears to decrease somewhat as a function of the intensity distortion. However, a Mann-Whitney *U* Test showed that this effect was not significant. This is in agreement with the lack of a significant difference in detection thresholds for intensity distortion between normal-hearing and hearing-impaired listeners (Sec. 5.II A; medians of the groups represented by arrows). The absence of a difference in sensitivity between normal-hearing and hearing-impaired listeners could be the result from the low perturbation levels used in this study. However, higher intensity distortion levels were not measured, because of unwanted spectro-temporal byproducts (see Chapter 4). In conclusion, the results do not show a relation between reduced speech intelligibility in noise and a distorted representation of intensity information.

- **Degradation of the temporal accuracy**

For all levels of temporal degradation, the medians of the SRTa's for the hearing-impaired listeners are higher than those for the normal-hearing listeners (Fig. 5.6b). The difference in performance between normal-hearing and hearing-impaired listeners does not decrease as a function of temporal perturbation. In addition, the group of hearing-impaired listeners performed as well as the normal-hearing listeners on the temporal perturbation detection task (Sec. 5.II A). In conclusion, the results do not suggest a relation between reduced intelligibility in noise and a distorted representation of temporal information.

- **Degradation of the spectral accuracy**

For the most extreme spectral perturbation condition, only the results using the male talker are used, because the male talker was just intelligible in this condition while the female talker was not (see Fig. 5.6c). The SRTa for the normal-hearing listeners in the spectral reference condition is about 3 dB higher than in the intensity and temporal reference condition, because the fine structure was perturbed in all spectral conditions (Sec. 5.I A). In the reference condition the median SRTa is higher for the hearing-

impaired listeners than for the normal-hearing listeners. When spectral perturbation is applied, the performance for the hearing-impaired listeners converges towards that for the normal-hearing listeners. At ¾-octave of spectral perturbation, the performance for the hearing-impaired listeners equals that for the normal-hearing listeners. Mann-Whitney *U* Tests confirm the observed trends: at 0 and ¼-octave perturbation the performance for the hearing-impaired listeners is significantly worse than that for the normal-hearing listeners ($p<0.05$), whereas at ½ and at ¾ octave no significant difference exists.

In summary, the detection threshold for spectral perturbation is significantly higher for hearing-impaired listeners than for normal-hearing listeners; moreover, convergence of the speech-processing performance of normal-hearing and hearing-impaired listeners is observed. This strongly points to a relation between a reduced intelligibility in noise and a distorted representation of spectral information.

D. Distortion-sensitivity model: Individual results

In the preceding section, the group results of the distortion-sensitivity model for normal-hearing and hearing-impaired listeners were compared. Now, the individual results will be used to further examine the relation between distorted coding of information and reduced speech intelligibility. As an estimate of individual performance, the sensitivity to the distortion was taken. The sensitivity to the distortion of individual listeners is defined as the slope of the linear regression line fitted through the individual SRTa values for different degrees of distortion. It quantifies how sensitive a listener is to the distortion of specific cues in speech. The underlying idea is that if a hearing-impaired listener is less sensitive to a particular artificial distortion than normal-hearing listeners, this artificially applied distortion probably relates to the internal deficit causing his speech perception problems. In this study two measures for suprathreshold speech perception quality are used: SII_{SRBT} and SII_{SRTa}. The relation between speech perception quality and the sensitivity to distortion of information will be evaluated.

For both intensity and temporal information, no correlation between the sensitivity to the distortion and SII_{SRTa} or SII_{SRBT} was observed in the individual data [Spearman rank correlation SII_{SRBT} and sensitivity to intensity distortion: -0.3 ($p=0.09$)].

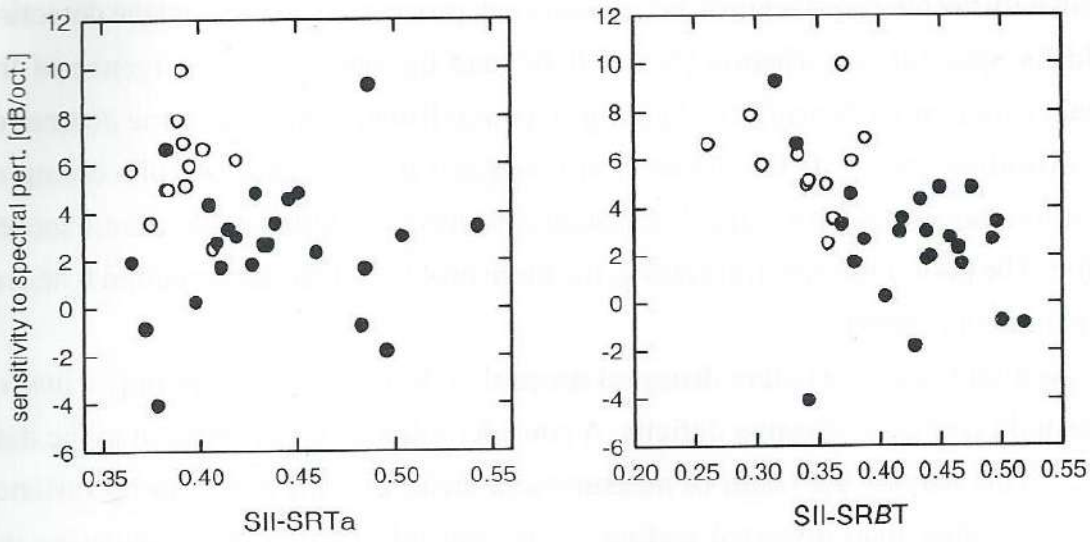


FIG. 5.7. The individual sensitivities to spectral perturbation for normal-hearing (open symbols) and hearing-impaired listeners (filled symbols) versus the SII_{SRTa} (panel a) and versus the SII_{SRBT} (panel b).

In Fig. 5.7, the sensitivity to distortion of spectral information is plotted against the individual SII_{SRTa} (panel a) and SII_{SRBT} (panel b). Open symbols represent the data for the normal-hearing listeners; filled symbols those for the hearing-impaired listeners. As is already clear from Fig. 5.6c, the median sensitivity of the hearing-impaired listeners is less than that of the normal-hearing listeners. No clear trend between SII_{SRTa} and sensitivity is shown [Spearman rank correlation: -0.2 (p=0.2)]; however, there is a correlation between SII_{SRBT} and sensitivity [Spearman rank correlation: -0.6 (p<0.05)]: the higher the SII_{SRBT}, the lower the sensitivity to spectral distortion.

SII_{SRTa} and SII_{SRBT} show a different picture: the sensitivity to spectral distortion is significantly correlated with the SII_{SRBT}, but not with the SII_{SRTa}. This difference may be explained by the different experimental setup: The Speech-Reception *Bandwidth* Threshold is measured using bandpass filtered speech signals embedded in complementary bandstop noise, whereas the speech-reception threshold test uses a noise spectrum equal to the average speech spectrum. Therefore, the SR*B*T is probably more sensitive to excessive spread of masking than the SRTa. As a result, the sensitivity to spectral distortion is likely to relate more directly to the SII_{SRBT} than to the SII_{SRTa}.

In summary, the individual results show a relation between suprathreshold speech processing as quantified by the SII_{SRBT} and the sensitivity to spectral distortion. This is in agreement with the observed relation between speech processing quality and the detection threshold for spectral perturbation (Sec. 5.II B), and the observed convergence of the performance for normal-hearing and hearing-impaired listeners for increasing degrees of spectral distortion (Sec. 5.II C). These results suggest that the auditory processing of spectral information of hearing-impaired listeners is distorted and that this affects speech perception. The poorer the spectral coding, the more problems hearing-impaired listeners have in perceiving speech.

The question remains whether distorted spectral auditory coding is the only cause of suprathreshold speech processing deficits. A considerable variance is present in the data of Fig. 5.7. This may be the result of measurement error, but this may also be variance due to factors other than distorted coding of spectral information. By calculating the reliability (Nunnally, 1967) of the variables in the correlation, an estimate of the influence of measurement error can be made. The square root of the product of the reliabilities of two tests gives an estimate of the unsigned maximum correlation possible, given the measurement accuracy.

The reliability of the SII_{SRTa} (6 measurements) is 0.9. The reliability of the sensitivity to the distortion is much smaller: about 0.3. This is because the measurement errors add up when the slope is estimated. Between SII_{SRTa} and sensitivity, the maximum unsigned correlation possible is about 0.5. The correlation observed was -0.2. Thus in the speech processing problems of hearing-impaired listeners as quantified by the SII_{SRTa}, spectral cues are probably not the only ones.

The reliability of the SII_{SRBT} (2 measurements) is 0.7. As a result, the estimate of the unsigned maximum correlation possible between SII_{SRBT} and sensitivity is 0.5. The correlation observed was -0.6. It may surprise that the absolute value of the observed correlation is larger than the predicted maximum correlation. However, the predicted maximum correlation is only a rough estimate. Therefore, all variance seems explained.

In summary, the distorted speech processing of hearing-impaired listeners measured by the SR*B*T test can fully be explained by distorted processing of spectral information, but with respect to the SRTa test other factors seem to affect intelligibility as well. This may be explained by the fact that upward spread of masking plays a dominant role in the SR*B*T test, but not in the SRTa test.

E. Comparison to literature

- **Degradation of the intensity accuracy**

The mean detection threshold for intensity distortion of hearing-impaired listeners is not significantly higher than that of normal-hearing listeners. However, some hearing-impaired listeners showed abnormally high distortion thresholds. This is consistent with the literature about intensity discrimination (for a review, see Florentine *et al.*, 1993). Overall, hearing-impaired listeners discriminate as well as normal-hearing listeners at equal sound pressure levels, and intensity discrimination may even be better at equal sensation levels. However, for some hearing-impaired listeners markedly higher discrimination thresholds are observed (Schroder *et al.*, 1994; Buus *et al.*, 1995).

With respect to speech intelligibility as a function of intensity distortion, no significant convergence of the performances for normal-hearing and hearing-impaired listeners was observed. In addition, no significant correlation between the sensitivity to intensity distortion and the SII was found. In contrast, in Chapter 4 a significant correlation between sensitivity to intensity distortion and SII_{SRBT} was observed. Several factors may account for this. Different listener groups were used in the previous and the present study. Since among hearing-impaired listeners a diversity of auditory deficits is observed (see, for example, Noordhoek *et al.*, submitted), this may lead to a different result. Moreover, although both groups of hearing-impaired listeners had comparable hearing loss, the presentation levels for the second group of listeners was 7 dB lower than for the first group due to lower uncomfortable loudness levels. Due to this difference in dynamic range, the same intensity perturbations may have introduced different loudness perturbations (see Chapter 4 of this thesis). These factors may explain why the correlation in the present study is not significant while in the previous study it was.

- **Degradation of the temporal accuracy**

The detection threshold for temporal distortion by hearing-impaired listeners was not significantly higher than that by normal-hearing listeners. However, some hearing-impaired listeners showed abnormally high detection thresholds. This is in agreement with the literature about temporal resolution. Although hearing-impaired listeners are known to suffer from excessive forward masking (Festen and Plomp, 1983; Oxenham and

Moore, 1995), on some tests of temporal resolution, most hearing-impaired listeners perform as well as normal-hearing listeners (Moore, 1995).

The performances for normal-hearing and hearing-impaired listeners did not converge as a function of the distortion of temporal information. In addition, no correlation between the sensitivity to temporal distortion and SII was observed. This agrees with the study of Duquesnoy and Plomp (1980). They measured how sensitive normal-hearing and hearing-impaired listeners were to reverberation. Reverberation can be considered a very systematic type of distortion of temporal information. Sensitivity was compared to the Speech Transmission Index (Houtgast and Steeneken, 1973). Their results showed that hearing-impaired listeners were as sensitive to reverberation as normal-hearing listeners.

- **Degradation of the spectral accuracy**

The detection thresholds for spectral distortion were significantly higher for the group of hearing-impaired listeners than for the group of normal-hearing listeners. In addition, convergence of speech perception performance for normal-hearing and hearing-impaired listeners as a function of spectral distortion was observed. This agrees with the results of Turner *et al.* (1995) that also showed convergence (see Introduction).

The results of this study suggest that hearing-impaired listeners suffer from reduced frequency selectivity and that this causes reduced speech intelligibility. This agrees with the literature, in which it has been reported frequently that hearing-impaired listeners suffer from reduced spectral resolution. [For review see Tyler (1986).] Reduced frequency selectivity affects speech intelligibility in two ways. First, because of reduced frequency selectivity the spectral contrasts in speech are less clear. Second, when frequency selectivity is reduced, hearing-impaired listeners will suffer from excessive upward and downward spread of masking.

Ter Keurs *et al.* (1992, 1993) investigated the first effect. Speech and noise, having the same long-term average spectrum, were added *after* the smearing of the spectral envelope. As a result, the effect of excessive masking was not simulated. Ter Keurs *et al.* (1993) observed that hearing-impaired listeners were as sensitive to reduced spectral contrasts in speech as normal-hearing listeners. They did find a small but significant correlation between the SRT for unsmeared speech and auditory filter bandwidth, but they could not explain this by a reduction of the spectral contrasts in speech.

In our study, the first and second effects were evaluated in combination, because first the noise was added to the speech and then the spectral information was distorted. Our results strongly suggest that reduced frequency selectivity influences speech intelligibility in noise. Since the results of ter Keurs *et al.* (1993) suggest that the first effect is not responsible for reduced speech perception, the reduced speech intelligibility in noise observed in hearing-impaired listeners is probably mainly due to the second effect, i.e., excessive spread of masking. Thus for hearing-impaired listeners, it is more difficult to separate speech from competing background noise.

III. SUMMARY AND CONCLUSIONS

In this study, the central question was how degraded speech perception of hearing-impaired listeners relates to distorted auditory coding. To investigate this, the intensity, time, and frequency information of sound were artificially distorted after wavelet coding. The detection thresholds for the different types of distortion were measured to obtain insight into how clearly hearing-impaired listeners could perceive a particular type of information. To investigate the relation between distorted auditory coding and speech perception, the distortion-sensitivity model was used. If hearing-impaired listeners are less sensitive with respect to speech perception than normal-hearing listeners to a particular type of distortion (intensity, time, or frequency), this indicates that this artificial distortion relates to the distorted auditory coding causing speech perception problems.

The group results showed that the detection thresholds for hearing-impaired listeners with respect to the distortion of intensity and temporal information were not significantly higher than those for normal-hearing listeners. For the distortion of spectral information, the detection thresholds for the hearing-impaired listeners were significantly higher than those for the normal-hearing listeners. Thus hearing-impaired listeners may perceive spectral information less clearly than normal-hearing listeners. With respect to the distortion-sensitivity model, the results (Fig. 5.6) showed that the group of hearing-impaired listeners was as sensitive as the group of normal-hearing listeners to intensity and temporal distortion. The group of hearing-impaired listeners was less sensitive than

normal-hearing listeners to the distortion of spectral information. Thus the group results suggest that distorted coding of spectral information is an important factor underlying the reduced speech intelligibility observed in hearing-impaired listeners.

Also, the individual results were considered to investigate the relation between reduced speech intelligibility and distorted coding of spectral information in more detail. A significant correlation between the SII, both SII_{SRTa} and $SII_{SR\mathit{B}T}$, and the detection threshold for spectral distortion was observed (Fig. 5.5). Thus the data reveal a statistical relation between the quality of speech processing, quantified by the SII, and the spectral coding accuracy, quantified by the detection threshold for spectral distortion. In addition, the correlation between the $SII_{SR\mathit{B}T}$ and the sensitivity to spectral distortion with respect to speech perception was significant (Fig. 5.7). Thus there is a statistical relation between the quality of speech processing and the effect of distortion of the spectral cues on speech perception. The more pronounced the speech perception problems of hearing-impaired listeners (in terms of the SII), the less accurate the spectral auditory coding (higher detection thresholds) and the less influence the distortion of spectral information has on speech intelligibility (lower sensitivity to spectral distortion). The individual results support the group result, strongly suggesting that distorted coding of spectral information is the factor underlying the suprathreshold problems encountered by many hearing-impaired listeners when trying to perceive speech.

The sensitivity to spectral distortion could explain all "true" variance in the $SII_{SR\mathit{B}T}$, i.e., all variance not due to measurement error. Thus distorted auditory coding of spectral information may be the only factor underlying speech processing deficits measured by means of the SR*B*T test. However, sensitivity to spectral distortion could not explain all "true" variance in the SII_{SRTa}. This suggests that, besides distorted coding of spectral information, other factors play a role in the suprathreshold speech processing problems of hearing-impaired listeners as reflected in the SRTa test.

From the data of the present study the following general conclusions can be drawn.

- The distortion-sensitivity model may be a valuable tool to investigate the underlying causes of reduced speech perception.
- Distorted auditory spectral coding may be an important factor underlying the speech perception problems of hearing-impaired listeners.
- Besides distorted coding of spectral information, other factors may play a role in reduced speech intelligibility as well.

REFERENCES

Allen, J. B. (**1977**). "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," IEEE Trans. Acoust. Speech Signal Process. **25**, 235–238.

Allen, J. B., and Rabiner, L. R. (**1977**). "A unified approach to short-time Fourier analysis and synthesis," Proc. of the IEEE **65**, 1558–1564.

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

Baer, T., and Moore, B. C. J. (**1993**). "Effects of spectral smearing on the intelligibility of sentences in noise," J. Acoust. Soc. Am. **94**, 1229–1240.

Bosman, A. J., and Smoorenburg, G. F. (**1995**). "Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing deficit," Audiology **34**, 260–284.

Buus, S., Florentine, M., and Zwicker, T. (**1995**). "Psychometric functions for level discrimination in cochlearly impaired and normal listeners with equivalent-threshold masking," J. Acoust. Soc. Am. **98**, 853–861.

Dreschler, W. A., and Plomp, R. (**1985**). "Relations between psychophysical data and speech perception for hearing-impaired subjects. II," J. Acoust. Soc. Am. **68**, 1261–1270.

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.

Festen, J. M., and Plomp, R. (**1983**). "Relations between auditory functions in impaired hearing," J. Acoust. Soc. Am. **73**, 652–662.

Florentine, M., Reed, C. M., Rabinowitz, W. M., Braida, L. D., and Durlach, N. I. (**1993**). "Intensity perception. XIV. Intensity discrimination in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **94**, 2575–2586.

Gabor, D. (**1947**). "Acoustical quanta and the theory of hearing," Nature (London) **159**, 591–594.

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. (**1995**). "Psycho-acoustics and speech recognition of the hearing impaired," in *Proceedings of the European Conference on Audiology*, Noordwijkerhout, The Netherlands, pp. 165–169.

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.

ISO (**1961**). International Organization for Standardization, ISO R226-1961, "Normal equal-loudness contours for pure tones and normal threshold of hearing under free field listening conditions" (available from American National Standards Institute, New York).

Moore, B. C. J. (**1995**), *Perceptual consequences of cochlear damage* (University Press: Oxford).

Moore, B. C. J. (**1996**), "Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids," Ear Hear. **17**, 133–161.

Moore, B. C. J., and Glasberg, B. R. (**1987**). "Relationship between psychophysical abilities and speech perception for subjects with unilateral and bilateral cochlear hearing impairments," in *The psychophysics of Speech Perception*, edited by M. E. H. Schouten (Nijhoff, Boston), pp. 449–460.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**1999**). "Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. I. Normal-hearing listeners," J. Acoust. Soc. Am. **105**, 2895–2902.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**in press**). "Measuring the threshold for speech-reception by adaptive variation of the signal bandwidth. II. Hearing-impaired listeners," to appear in J. Acoust. Soc. Am..

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**submitted**). "Speech perception related to auditory functions at 1 kHz for hearing-impaired listeners," submitted to J. Acoust. Soc. Am..

Nunnally, J. C. (**1967**). *Psychometric theory* (McGraw-Hill, New York), pp. 172–235.

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

Patterson, R. D., and Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788–1803.

Pavlovic, C. V. (**1987**). "Derivation of primary parameters and procedures for use in speech intelligibility predictions," J. Acoust. Soc. Am. **82**, 413–422.

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the Speech Reception Threshold for sentences," Audiology **18**, 43–52.

Rioul, O., and Vetterli, M. (**1991**). "Wavelets and signal processing," IEEE Signal Proc. Mag. October, 14–38.

Scharf, B. (**1970**). "Critical bands," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. 1, pp. 157–202.

Schroder, A. C., Viemeister, N. F., and Nelson, D. A. (**1994**). "Intensity discrimination in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **96**, 2683–2693.

Smoorenburg, G. F. (**1992**). "Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram," J. Acoust. Soc. Am. **91**, 421–437.

ter Keurs, M., Festen, J. M., and Plomp, R. (**1992**). "Effect of spectral envelope smearing on speech reception. I.," J. Acoust. Soc. Am. **91**, 2872–2880.

ter Keurs, M., Festen, J. M., and Plomp, R. (**1993**). "Limited resolution of spectral contrast and hearing loss for speech in noise," J. Acoust. Soc. Am. **94**, 1307–1314.

Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). "Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 2568–2576.

Tyler, R. S. (**1986**). "Frequency resolution in hearing-impaired listeners," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic Press, London), pp. 309–371.

Tyler, R. S., Summerfield, Q., Wood, E. J., and Fernandes, M. A. (**1982**). "Psychoacoustic and temporal processing in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **72**, 740–752.

van Rooij, J. C. G. M., and Plomp, R. (**1990**). "Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses," J. Acoust. Soc. Am. **88**, 2611–2624.

van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**1999**). "Intensity discrimination of Gaussian-windowed tones: indications for the shape of the auditory frequency-time Window," J. Acoust. Soc. Am. **105**, 3425–3435.

van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**submitted**). "The effect of intensity perturbations on speech intelligibility for normal-hearing and hearing-impaired listeners," submitted to J. Acoust. Soc. Am.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**in press**) "Method for the selection of speech materials for efficient measurement of the speech reception threshold," to appear in J. Acoust. Soc. Am..

Yantis, P. A. (**1994**). "Puretone air-conduction threshold testing," in *Handbook of Clinical Audiology*, edited by J. Katz (Williams and Wilkins, Baltimore, USA), 4th ed., Chap. 7, p. 106.

6 General discussion

After using wavelet coding of sound as a tool to study the auditory system, the usefulness of this tool will be discussed in this final chapter. A review will be given of the results yielded using a wavelet coding tool tailored to the auditory system.

In this thesis, wavelet coding is used as a tool to study the auditory system. This strategy was chosen because of an important similarity between wavelet coding and auditory coding, namely that their spectral resolutions are constant on a logarithmic frequency axis. Investigating the auditory system by means of wavelet coding seemed interesting, and this final chapter gives a review of the results of this wavelet approach. First, the similarities between auditory coding and wavelet coding will be discussed (with respect to part I of this thesis). Then, the results obtained by using auditory wavelet coding will be discussed (with respect to part II of this thesis). Subsequently, possible directions for further investigation will be given. Finally, the success of wavelet coding as a tool for studying the auditory system will be considered.

I. PART I: AUDITORY CODING AND WAVELET CODING

In the first part of this thesis, the auditory time-frequency window, i.e., the window with which the ear analyzes sound, was characterized. In Chapter 2, intensity discrimination experiments were performed with Gaussian tone pulses and the results were interpreted in terms of the multiple-window model: intensity discrimination improves when more auditory windows are involved in the perception. The hypothesis was that, when measuring intensity discrimination as a function of spectro-temporal shape, the intensity discrimination performance is worst when the spectro-temporal shape of the stimulus matches the spectro-temporal shape of the auditory window most closely. As a result, the just-noticeable differences in intensity were expected to show a convex shape when plotted as a function of the spectro-temporal shape. Indeed, this convex shape ("hump") was observed. The results measured at carrier frequencies of 1 kHz and 4 kHz were similar: both implied a corresponding bandwidth of the stimulus between roughly ¼ and ⅓ octave. Since the stimuli used were Gaussian tone pulses for which the durations are inversely proportional to the bandwidth, the corresponding durations of these stimuli were about 4 ms at 1 kHz and 1 ms at 4 kHz. The similarity of the intensity discrimination performances at 1 and 4 kHz plotted as functions of the spectro-temporal shape suggests that also for the auditory window the duration is roughly inversely proportional to the bandwidth: if the duration of the auditory window at 4 kHz was larger than 1 ms, the "hump" in the intensity discrimination performance would be broader. This was not observed. Thus the spectral width of the auditory time-frequency window increases with increasing frequency while the temporal width decreases.

In psychoacoustics, it is generally accepted that spectral resolution decreases with increasing frequency: the auditory periphery can be though of as a bank of bandpass filters, each between ¼ and ⅓ octave wide, related to the auditory critical band (see Scharf, 1970). The spectral widths of the auditory time-frequency windows as determined in this thesis are in close agreement with these critical bands.

More controversial is the duration of the auditory time-frequency window. Our conclusion that the duration decreases with increasing frequencies is in general agreement

with results in the literature (Florentine *et al.*, 1988; Gerken *et al.*, 1990; Plack and Moore, 1990). However, in absolute terms the temporal resolution observed in these studies is usually larger than our result of a few milliseconds. For example, Plack and Moore (1990) found durations of 13 ms at 0.3 kHz decreasing to 7 ms at 8.1 kHz. However, in a recent study by Wiegrebe and Krumbholz (1999), temporal resolution was almost a factor 10 higher. Wiegrebe and Krumbholz argue that different parts of the auditory pathway will limit temporal resolution measured in different experiments. In their experiments and in the experiments of Chapter 2 of this thesis, the temporal resolution of the auditory periphery was probably the limiting factor. On the contrary, in the gap detection experiments of Plack and Moore (1990), more central parts of the auditory system probably limited performance. Thus, in their experiments, the temporal information may be available in the auditory nerve, but the central auditory system was not able to use it optimally for gap detection.

With respect to the multiple-window model used to explain the intensity discrimination results, Baer *et al.* (1999a) questioned the validity of this model. They reproduced the results of Chapter 2 and explained them using an alternative theory based on basilar membrane compression. Baer *et al.* state that the improvement of intensity discrimination for short-duration broadband clicks is not due to the combination of information of different auditory time-frequency windows (multiple-window model), but due to the input-output function on the basilar membrane being less compressive for very brief stimuli during the initial part of the response. Although not mentioned by Baer *et al.*, it seems likely that for these short-duration broadband stimuli the information from different auditory filters (windows) is combined (Florentine and Buus, 1981). Therefore, both compression and multiple windows may play a role in intensity discrimination. More experiments will be needed to test the two hypotheses.

In summary, in the first part of this thesis an attempt was made to characterize time-frequency analysis of the peripheral auditory system. Based on the results of intensity discrimination experiments using Gaussian tone pulses, the bandwidth of the auditory time-frequency window was estimated at about ¼ octave. This is in agreement with the classical view of spectral processing of the auditory periphery, i.e., a bank of bandpass filters each a critical band wide (about ¼ octave). In many models of the auditory system, spectral and temporal processing are considered independently. A bandpass filter has a certain time constant, but estimating the time constant of the auditory periphery from the

filter bandwidth is tricky, unless additional assumptions about the order of the filter are made (de Boer, 1985). In this thesis the time constant was coupled to the frequency constant by estimating the time-frequency shape of a Gaussian tone pulse that matches the auditory time-frequency window best. The spectral width of this function is proportional to frequency and the temporal window inversely proportional to frequency. A time-frequency analysis with these characteristics can be considered a wavelet analysis. Therefore, in Chapter 3 of this thesis a wavelet analysis and synthesis method was developed as an attempt to mimic time-frequency coding of the auditory periphery.

II. PART II: SPEECH PERCEPTION AND DISTORTED CODING

In part II of this thesis, the effects of distorted auditory coding on speech perception were examined, using a wavelet decomposition and recomposition scheme as a signal-processing tool. The auditory wavelet coding was used to model normal auditory time-frequency coding. By distorting the wavelet coefficients, distorted auditory coding was simulated. Intensity, temporal, or spectral information of speech was distorted and the effect on speech perception was measured. The distortion-sensitivity model was used, comparing the results of hearing-impaired listeners on speech perception as a function of artificial distortion with the results of normal-hearing listeners. The underlying idea of this approach is that if a hearing-impaired listener is less sensitive to a particular type of distortion than normal-hearing listeners, this artificial distortion relates to the hearing-impaired listener's distorted (suprathreshold) auditory coding that degrades his speech perception performance.

Distortion of the intensity information of sound was obtained by multiplying the modulus of each wavelet coefficient by a random factor. Temporal and spectral information was distorted by randomly shifting the position of each wavelet along the temporal and spectral axis, respectively. The effect of distortion of one dimension on the information content of other dimensions was taken into account. This method for simulating a distorted representation of information was chosen because of the elegant

possibility to treat the three dimensions in a way that is essentially identical. This emphasizes the intrinsic link between intensity, time and frequency. Still, this study did not aim at exactly simulating possible coding deficiencies in the auditory system, but the essential effects of the auditory coding deficits should be simulated. The method appears to be able to do this satisfactorily, at least with respect to distorted auditory coding of spectral information.

Chapter 5 showed that main effects were observed for spectral coding only: The detection thresholds for the artificial distortion of spectral information in a group of hearing-impaired listeners were higher than those of normal-hearing listeners. In addition, hearing-impaired listeners were less sensitive than normal-hearing listeners to artificial spectral distortion when trying to understand speech in noise. This strongly suggests that hearing-impaired listeners suffer from distorted auditory coding of spectral information and that this causes problems in speech perception. There were also indications that other factors, besides reduced spectral resolution, limited speech perception. The results of Chapter 4 suggest that distorted coding of intensity information may play a role. However, this could not be concluded from the results of Chapter 5. The results in this chapter did not show a link between distorted coding of temporal information and reduced speech intelligibility.

During this project, speech perception of hearing-impaired listeners was studied in the same group using a different approach, i.e., the correlation approach combined with the examination of individual data ("individual approach") (Noordhoek *et al.*, submitted). In a correlation approach, the individual differences among listeners are used to study statistical relations between auditory functions, e.g., frequency selectivity, and speech perception. In the "individual approach," the individual results are looked at, to examine whether a hearing-impaired listener who performed less with respect to speech intelligibility was also performing poorly with respect to one or more auditory functions. Noordhoek *et al.* measured the auditory functions around 1 kHz. When measuring speech intelligibility, the bandwidth of the speech was limited to a frequency region around 1 kHz to such an extent that intelligibility of short sentences dropped to 50% (Speech-Reception *Bandwidth* Threshold test). Noordhoek *et al.* accounted for the effect of hearing threshold and sound pressure level of the stimuli on speech perception by means of the Speech Intelligibility Index as was followed by the present study.

The main conclusions of Noordhoek *et al.* (submitted) agree with those of this thesis: reduced spectral resolution is a very important cause for speech perception deficits. Both in this study and in the study by Noordhoek *et al.*, it became clear that factors other than reduced spectral resolution and hearing threshold can also affect speech perception of hearing-impaired listeners. The group of hearing-impaired listeners of Noordhoek *et al.* did not appear to suffer from distorted coding of intensity information, as they did not perform worse than normal-hearing listeners in the intensity discrimination task. In contrast, the results of Chapter 4 of this thesis suggest that distorted intensity coding may play a role. Noordhoek *et al.* showed that reduced temporal resolution may be a factor underlying reduced speech intelligibility for some hearing-impaired listeners. In contrast, in Chapter 5 of this thesis, distorted coding of temporal information did not seem to be a factor. To summarize, the importance of good acuity of spectral coding for speech perception became clear both in this study and in the study by Noordhoek *et al.*. The role of intensity and temporal coding is still less clear.

Both studies showed that reduced spectral resolution is an important cause for speech perception problems. However, they also showed that large differences among hearing-impaired listeners exist, with respect to the seriousness of the deficit, but also with respect to the type of auditory deficit. The studies make clear that, to be able to help hearing-impaired listeners, it is important to examine their individual auditory deficits and how they affect speech perception. After "earmarking" a hearing-impaired listener's problem, this listener could benefit from an individual correction.

III. SUGGESTIONS FOR FURTHER STUDIES

More research is needed to clarify the role of intensity and temporal coding for speech perception. The strongest approach is probably an approach from two directions: the correlation/individual approach, and the distortion-sensitivity model. The correlation/individual approach can show the auditory deficits from which hearing-impaired listeners are suffering, and how they correlate with speech intelligibility. Then, by the distortion-sensitivity model, the effect of these auditory deficits on speech

perception can be examined more directly. Measures of temporal auditory deficits are forward masking, backward masking, and amplitude modulation detection (temporal modulation transfer function). For slow amplitude modulation rates, modulation detection can also be seen as an example of a measure of intensity processing, like intensity discrimination. Applying the distortion-sensitivity model, speech perception as a function of distortion of intensity or temporal information of hearing-impaired listeners is compared with that of normal-hearing listeners. It might prove useful to examine the effects of various types of distortions. For example, in a study on the acuity of intensity coding, one could also use sparse coding of the wavelet coefficients instead of multiplying the modulus of each wavelet coefficient by a random factor. The idea behind the intensity perturbation of Chapters 4 and 5 was to simulate reduced acuity of auditory intensity coding due to noisy intensity information. Sparse coding would simulate another aspect of reduced acuity of auditory intensity coding, i.e., a coding with low intensity "selectivity." It can also be used to study impaired loudness perception.

To study the relation between speech perception and the processing of temporal information, it might be worthwhile to use fluctuating noise maskers besides the nonfluctuating noise maskers. Effects of distorted coding of temporal information are probably more pronounced when fluctuating noise maskers are used than when unmodulated maskers are used.

In this study a correction to the speech perception values with respect to effects of audibility was applied by means of the Speech Intelligibility Index model. The SII model is based on the spectra of speech and maskers. It can estimate the effect of a continuous steady-state noise masker on speech intelligibility. However, it cannot deal with fluctuating maskers. Therefore, in a further study of processing of temporal information, the SII model should be extended to include fluctuating noise sources as well, for example by means of the phase-locked modulation transfer function (Ludvigsen *et al.*, 1990; Drullman *et al.*, 1994, 1996).

With respect to the shape of the auditory time-frequency window, it is interesting to investigate the shape of this window for hearing-impaired listeners. In this way, it may be possible to test whether the compression model proposed by Baer *et al.* (1999a) or our multiple-window model is better suited to explain the intensity discrimination experiments for stimuli with different spectro-temporal shapes. The model of Baer *et al.* explains the variation in intensity discrimination by a variation in the degree of amplitude

compression. Since hearing-impaired listeners experience less amplitude compression than normal-hearing listeners, their model predicts that intensity discrimination for hearing-impaired listeners varies less as a function of spectro-temporal shape than that for normal-hearing listeners. In contrast, the multiple-window model predicts that intensity discrimination will be worst for stimuli with spectro-temporal shapes that correspond most closely to those of the auditory time-frequency windows. Since the auditory time-frequency window of hearing-impaired listeners is expected to have an increased bandwidth, the poorest discrimination performance is expected for stimuli with a broader bandwidth than that observed for normal-hearing listeners.

Recently, Baer *et al.* (1999b) measured intensity discrimination as a function of spectro-temporal shape for hearing-impaired listeners. Although large variability was present in the results, they seem to provide support for both hypotheses: for some hearing-impaired listeners, intensity discrimination did not vary as a function of spectro-temporal shape; for other hearing-impaired listeners, the intensity-discrimination results showed a "hump" for stimuli with a broader bandwidth than that observed for the normal-hearing listeners. These results suggest that probably both mechanisms apply to some extent in intensity discrimination.

IV. WAVELET CODING AS A TOOL FOR STUDYING THE AUDITORY SYSTEM?

The underlying reason for using wavelet coding to study the auditory system was the presumption that peripheral auditory time-frequency coding is very similar to wavelet coding. The results of the first part of this thesis confirm this. Just like for a wavelet, the bandwidth of the auditory window is (roughly) proportional to frequency and the duration of the window is (roughly) inversely proportional to frequency. Therefore, modeling auditory spectro-temporal coding by wavelet coding seems highly appropriate. In the second part of this thesis, distorted auditory coding was mimicked by a distortion of the wavelet coding. Hearing-impaired listeners were less sensitive to the distortion of spectral information than normal-hearing listeners, and this strongly suggests that a distorted

representation of spectral information is the cause for reduced (suprathreshold) speech perception of hearing-impaired listeners.

Let us finally return to the point raised at the beginning of this chapter, i.e., the successfulness of wavelet coding as a tool to study the auditory system. It was shown that, in comparison with short-time Fourier analysis, wavelet analysis simulates the time-frequency analysis of the auditory system more closely with respect to temporal and spectral resolution. Therefore, it is worthwhile to consider wavelet coding in studies of the auditory system. Wavelet coding is an interesting and easy manageable tool for further investigation. In this thesis, this tool provided some insight into the suprathreshold speech processing problems of hearing-impaired listeners: these problems mainly result from distorted auditory processing of spectral information.

REFERENCES

Baer, T., Moore, B. C. J., and Glasberg, B. R. (**1999a**). "Detection and intensity discrimination of Gaussian-shaped tone pulses as a function of duration," J. Acoust. Soc. Am. **106**, 1907–1916.

Baer, T., Marriage, J., and Moore, B. C. J. (**1999b**). "Intensity discrimination of brief gaussian-windowed tones by the hearing impaired," BSA short papers, Essex, 34.

de Boer, E. (**1985**). "Auditory time constants: A paradox?," in *Time Resolutions of Auditory Systems*, Proceedings of the 11th Danavox Symposium on Hearing, Gamle Avernæs, Denmark, pp. 141–158.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of reducing slow temporal modulations on speech reception," J. Acoust. Soc. Am. **95**, 2670–2680.

Drullman, R., Festen J. M., and Houtgast T. (**1996**). "Effect of temporal modulation reduction on spectral contrasts in speech," J. Acoust. Soc. Am. **99**, 2358–2364.

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

Florentine, M., Fastl, H., and Buus, S. (**1988**). "Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking," J. Acoust. Soc. Am. **84**, 195–203.

Gerken, G. M., Bhat, V. K. H., and Hutchison-Clutter, M. H. (**1990**). "Auditory temporal integration and the power-function model," J. Acoust. Soc. Am. **88**, 767–778.

Ludvigsen, C., Elberling, C., Keidser, G., and Poulsen, T. (**1990**). "Prediction of intelligibility of non-linearly processed speech," Acta Otolaryngol. Suppl. **469**, 190–195.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**submitted**). "Speech perception related to auditory functions at 1 kHz for hearing-impaired listeners," submitted to J. Acoust. Soc. Am..

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

Scharf, B. (**1970**). "Critical bands," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. 1, pp. 157–202.

Wiegrebe, L., and Krumbholz, K. (**1999**). "Temporal resolution and temporal masking properties of transient stimuli: Data and an auditory model," J. Acoust. Soc. Am. **105**, 2746–2756.

Summary: Wavelet coding of sound as a tool for studying the auditory system

In this thesis, wavelet coding is used as a tool for studying the time-frequency behavior of the auditory system. The reason for using wavelet analysis, instead of the often-used short-time Fourier analysis is the resemblance between wavelet analysis and auditory analysis with respect to spectral resolution. In short-time Fourier analysis, spectral resolution is constant throughout the frequency scale. On the contrary, in wavelet analysis, spectral resolution is proportional to frequency. For the auditory system, above about 500 Hz, spectral resolution is proportional to frequency as well. Therefore, a wavelet coding algorithm is developed that mimics the time-frequency analysis of the auditory system. With the aid of an intensity-discrimination experiment, the temporal and spectral resolution of the wavelet coding are tuned to the resolution of the auditory system. Then, the resulting "auditory" wavelet coding is used as a front-end signal-processing tool in studying the auditory system. Artificial distortion of this wavelet coding is used to simulate the effects of distorted auditory coding on speech perception.

This thesis starts with an abstract consideration of auditory time-frequency analysis. To analyze sounds of different time-frequency shapes, the auditory system performs a time-frequency analysis using time-frequency windows. A stimulus can give excitation in a single or in several time-frequency windows, depending on its time-frequency shape. When more than one time-frequency window is excited by a stimulus, it is reasonable to assume that, in an intensity discrimination task, the information from the different windows is combined statistically. This is described in the so-called multiple-window model, which is a generalization of the multiband excitation pattern model (Florentine and Buus, 1981) in which the information from different auditory filters is combined, and

the multiple-look model (Viemeister and Wakefield, 1991) in which the information from different time segments or looks is combined.

In Chapter 2 of this thesis, the multiple-window model is tested on its merits in an intensity discrimination experiment using Gaussian tone pulses. The time-frequency shape of these stimuli is varied from a long-duration narrow-band tone to a short-duration broadband click. Since these stimuli have different time-frequency shapes, different numbers of windows may be excited: a series of windows along the temporal axis for the long-duration narrow-band tone; a series of windows along the spectral axis for the short-duration broadband click; and one or only a few windows for an intermediate tone pulse. The multiple-window model predicts that the more windows are involved in the intensity discrimination task, the better the performance will be, because information from different windows can be processed independently and combined subsequently; performance will be poorest for a stimulus of intermediate shape that excites one or only a few windows.

The intensity discrimination results fit well into the multiple-window model. Intensity discrimination performance as a function of time-frequency shape has a "convex" shape, with poorest performance for a stimulus with a bandwidth of about ¼ octave. This "critical" bandwidth is observed both at 1 kHz and at 4 kHz, suggesting that the bandwidth of the auditory window is proportional to frequency. As a consequence of the use of Gaussian tone pulses, the durations of the "critical" stimuli are inversely proportional to frequency. The similar results at 1 and 4 kHz suggest that the duration of the auditory time-frequency window is inversely proportional to frequency as well, because the width of the "convex" shape would be different for different frequencies, if the duration of the auditory window was not inversely proportional to frequency. In summary, the results from this experiment indicate that the spectral width of the auditory window is proportional to frequency and the temporal width inversely proportional to frequency. Thus peripheral auditory time-frequency coding can be approximated by wavelet coding.

The intensity discrimination results suggest that a Gaussian mother wavelet, i.e., a complex sinusoidal carrier with a Gaussian envelope, with a bandwidth of ¼ octave approximates the auditory time-frequency window. Using this mother wavelet, a decomposition and recomposition method is developed, as described in Chapter 3. Nyquist's sampling theorem is used to decide on an adequate sampling in time and

frequency (Allen, 1977; Allen and Rabiner, 1977). The resulting time-frequency sampling is eight wavelets per octave along the spectral axis, and one wavelet per three stimulus periods along the temporal axis. This "auditory" wavelet coding tool is used as a signal-processing tool in subsequent studies of the auditory system.

In Chapters 4 and 5, the developed "auditory" wavelet coding is used to study speech perception of hearing-impaired listeners. Many hearing-impaired listeners have problems to understand speech in noise, even if sounds are well above the hearing threshold. These listeners possibly suffer from a distorted auditory coding. The effect of this distorted auditory coding on speech perception is studied by artificially distorting the wavelet coding between decomposition and recomposition of sound, and measuring the effect of this artificial distortion on speech intelligibility. Perturbations are applied in three dimensions of coding: intensity, time, and frequency. The effects of distorted coding in each of these dimensions are interpreted using the so-called distortion-sensitivity model. In this model, speech perception performance as a function of the degree of distortion is compared between hearing-impaired listeners and normal-hearing listeners. The underlying idea is that, when the auditory coding of a particular cue in sound is distorted for hearing-impaired listeners, they will be less sensitive to an artificial distortion of that cue than normal-hearing listeners. If speech perception of hearing-impaired listeners is affected less by the distortion than that of normal-hearing listeners, performance of normal-hearing and hearing-impaired listeners will converge towards higher degrees of distortion. Thus, convergence for a particular type of distortion is an indication that this artificial distortion relates to the auditory distorted coding that causes speech perception problems.

The results of Chapter 5 did not show that hearing-impaired listeners were less sensitive to a distorted coding of intensity or temporal information than normal-hearing listeners. This suggests that auditory coding in the intensity or the temporal domain does not constitute a problem for hearing-impaired listeners, although in Chapter 4 some indications that auditory intensity coding might be a problem were given. On the other hand, with regard to spectral perturbations, speech perception performance for hearing-impaired listeners is clearly less sensitive than performance for normal-hearing listeners. In addition, hearing-impaired listeners also had problems to detect such spectral distortions. The low sensitivity of hearing-impaired listeners with respect to spectral

distortion suggests that their problems in understanding suprathreshold speech in noise are due to coding problems with respect to spectral information.

In conclusion, wavelet coding approximates peripheral auditory coding, as confirmed by the intensity discrimination experiments using Gaussian tone pulses. From the results of this experiment, a wavelet coding algorithm, using a Gaussian mother wavelet with a bandwidth of ¼ octave, is developed to model auditory time-frequency coding. Artificial distortion of the wavelet coding is used to investigate the influence of distorted auditory coding on speech perception. The results of this study suggest that distorted auditory coding of spectral information is an important factor underlying speech perception problems of hearing-impaired listeners.

REFERENCES

Allen, J. B. (1977). "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," IEEE Trans. Acoust. Speech Signal Process. **25**, 235–238.

Allen, J. B., and Rabiner, L. R. (**1977**). "A unified approach to short-time Fourier analysis and synthesis," Proc. of the IEEE **65**, 1558–1564.

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

Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, Pt. 1, 858–865.

Samenvatting: Waveletcodering van geluid als middel voor het bestuderen van het auditief systeem

In dit proefschrift wordt waveletcodering gebruikt om de tijd-frequentie analyse van het gehoor te bestuderen. De reden voor het gebruik van een waveletcodering in plaats van de klassieke *short-time* Fouriertransformatie is de overeenkomst wat betreft spectrale resolutie van waveletanalyse en auditieve tijd-frequentie analyse. Bij een *short-time* Fouriertransformatie is de spectrale resolutie constant over de gehele frequentieschaal. Echter, bij een waveletanalyse is de spectrale resolutie evenredig met de frequentie, evenals bij het auditief systeem (bij frequenties hoger dan ongeveer 500 Hz). Met behulp van intensiteitsdiscriminatie-experimenten is bepaald hoe de temporele en spectrale resolutie van de waveletanalyse kan worden aangepast aan die van het gehoor. Vervolgens is met behulp van een kunstmatige verstoring van de waveletcodering onderzocht wat het belang is van verschillende aspecten van perifere auditieve codering voor spraakperceptie.

Dit proefschrift begint met een poging om de tijd-frequentie analyse van de auditieve periferie te karakteriseren. Voor het verwerken van zowel temporele als spectrale variaties in geluid voert het gehoor een tijd-frequentie analyse uit met behulp van tijd-frequentie vensters (*windows*). Een stimulus kan één of meerdere auditieve vensters activeren afhankelijk van zijn vorm. Wanneer er meer dan één tijd-frequentie venster wordt geactiveerd is het redelijk om aan te nemen dat bij een discriminatietaak de statistische informatie van verschillende vensters wordt gecombineerd. Dit wordt in dit proefschrift het *multiple-window model* genoemd. Het is een combinatie van het *multiband excitation pattern model* (Florentine en Buus, 1981) waarin de informatie van verschillende auditieve filters wordt gecombineerd, en het *multiple-look model* (Viemeister en

Wakefield, 1991) waarin de informatie van verschillende tijdsegmenten (*looks*) wordt gecombineerd.

In hoofdstuk 2 van dit proefschrift wordt het *multiple-window model* getest in een intensiteitsdiscriminatie-experiment met Gaussische toonpulsen. De spectro-temporele vorm van deze stimuli is gevarieerd van een toon (lange duur, smalbandig) naar een klik (korte duur, breedbandig). Omdat deze stimuli verschillen wat betreft hun spectro-temporele vorm, zullen verschillende stimuli een verschillend aantal tijd-frequentie vensters aanslaan: een aantal vensters langs de tijd-as door een toon; een aantal vensters langs de frequentie-as door een klik; één (of een paar) vensters door een stimulus die zich wat betreft spectro-temporele vorm tussen toon en klik in bevindt. Het *multiple-window model* voorspelt dat, hoe meer vensters betrokken zijn bij de intensiteitsdiscriminatietaak, hoe makkelijker de beslissing zal zijn, omdat de informatie van verschillende vensters onafhankelijk verwerkt kan worden en vervolgens kan worden gecombineerd; de beslissing zal het moeilijkst zijn voor een stimulus met een tussenvorm die slechts één of een paar vensters aanslaat.

De resultaten van de intensiteitsdiscriminatie-experimenten komen overeen met de voorspellingen van het *multiple-window model*. De taak is inderdaad het moeilijkst voor een stimulus met een spectro-temporele vorm tussen toon en klik. Deze stimulus heeft een bandbreedte van ongeveer ¼ octaaf. Deze "kritische" bandbreedte wordt gemeten bij zowel 1 kHz als bij 4 kHz. Dit toont aan dat, in ieder geval bij de gemeten frequenties, de bandbreedte van het auditieve venster evenredig is met frequentie. Omdat bij de experimenten gemeten is met Gaussische toonpulsen, is de duur van de "kritische" stimuli omgekeerd evenredig met de frequentie. Echter, de gelijkvormige resultaten bij 1 en 4 kHz suggereren dat de duur van het auditieve tijd-frequentie venster ook omgekeerd evenredig is met de frequentie. Immers, de breedte van het maximum zou verschillend zijn bij verschillende frequenties als de duur van het auditieve venster niet omgekeerd evenredig met de frequentie zou zijn. Dus, waveletcodering lijkt een goede benadering voor de tijd-frequentie codering van de auditieve periferie.

De intensiteitsdiscriminatie-experimenten laten zien dat de auditieve tijd-frequentie analyse kan worden benaderd met een Gaussisch moederwavelet (een complexe sinusvormige draaggolf met een Gaussische omhullende) met een bandbreedte van ¼ octaaf. Met dit moederwavelet is een waveletanalyse en -synthese ontwikkeld, die beschreven is in hoofdstuk 3. Nyquist's bemonsteringstheorema is gebruikt om de

bemonstering in tijd en frequentie te bepalen (Allen, 1977; Allen en Rabiner, 1977). De resulterende tijd-frequentie bemonstering is één wavelet per drie waveletperiodes langs de tijd-as en acht wavelets per octaaf langs de frequentie-as. Deze waveletcodering is vervolgens gebruikt om het gehoor te bestuderen.

In hoofdstuk 4 en 5 wordt beschreven hoe met de ontwikkelde 'auditieve' waveletcodering de problemen met spraakverstaan van slechthorenden zijn onderzocht. Veel slechthorenden hebben problemen met het verstaan van spraak in rumoer, zelfs als het geluid boven de gehoordrempel is. Deze luisteraars hebben blijkbaar last van een verstoorde auditieve codering. In dit proefschrift is het effect van een verstoorde auditieve codering op het spraakverstaan onderzocht door een kunstmatige verstoring van de waveletcodering. De verstoringen zijn aangebracht in drie dimensies: intensiteit, tijd en frequentie. Spraakverstaan is gemeten als functie van de mate van verstoring voor normaal- en slechthorenden en geïnterpreteerd met het *distortion-sensitivity model*. De achterliggende gedachte van dit model is dat, indien een verstoorde verwerking van bepaalde informatie een oorzaak is van problemen met spraakverstaan, slechthorenden minder dan normaalhorenden last zullen hebben van een kunstmatige verstoring van deze informatie. In dat geval zullen de prestaties van normaal- en slechthorenden naar elkaar toe groeien als functie van de mate van verstoring. Met andere woorden, convergentie voor een bepaald type verstoring is een aanwijzing dat deze kunstmatige verstoring een relatie heeft tot de verstoorde auditieve codering waar het spraakverstaan van slechthorenden onder lijdt.

De resultaten van hoofdstuk 4 suggereren dat een verstoorde auditieve intensiteitscodering een mogelijke oorzaak van problemen met spraakverstaan is. Echter, de resultaten van hoofdstuk 5 bevestigen dit niet. Voor een verstoring van temporele informatie waren slechthorenden ook niet minder gevoelig dan normaalhorenden. Daarmee toonde het onderzoek niet aan dat de auditieve codering van intensiteit- of temporele informatie een probleem is voor slechthorenden. Het spraakverstaan van slechthorenden leed duidelijk wel minder onder een kunstmatige verstoring van spectrale informatie dan dat van normaalhorenden. Bovendien hadden slechthorenden meer moeite met het detecteren van de spectrale verstoring. Dit toont aan dat problemen met het coderen van spectrale informatie een oorzaak zijn van de problemen van slechthorenden met het verstaan van spraak in rumoer. Uit het onderzoek bleek verder dat er onder slechthorenden grote onderlinge verschillen bestaan in de problemen met spraakverstaan

en dat er, naast een slechte verwerking van frequentie-informatie, waarschijnlijk ook ander factoren een rol spelen.

Samenvattend, de tijd-frequentie analyse van de auditieve periferie kan worden benaderd met een waveletcodering. Resultaten van intensiteitsdiscriminatie-experimenten met Gaussische toonpulsen maken dat aannemelijk. Met behulp van deze resultaten is een waveletanalyse ontwikkeld, gebruik makend van een Gaussisch moederwavelet met een bandbreedte van ¼ octaaf. Een kunstmatige verstoring van de waveletcodering is gebruikt om de effecten van een verstoorde auditieve codering op het spraakverstaan te simuleren. De resultaten laten zien dat een slechte verwerking van spectrale informatie een oorzaak is van de bovendrempelige problemen met spraakverstaan waar slechthorenden onder lijden.

REFERENTIES

Allen, J. B. (**1977**). "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," IEEE Trans. Acoust. Speech Signal Process. **25**, 235–238.

Allen, J. B., en Rabiner, L. R. (**1977**). "A unified approach to short-time Fourier analysis and synthesis," Proc. of the IEEE **65**, 1558–1564.

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

Viemeister, N. F., en Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, Pt. 1, 858–865.

Bedankt

Een promotie wordt wel vergeleken met het beklimmen van een berg. Ik vind dit een mooie vergelijking. Mijn promotie bevatte steile stukjes, maar juist die maakten het de moeite waard. Aan de vergelijking wordt soms ook nog toegevoegd dat het niet meevalt om boven te komen, in je eentje. Gelukkig stond ik er niet alleen voor. Hierbij bedank ik iedereen die een bijdrage heeft geleverd aan mijn promotieonderzoek.

In de eerste plaats mijn promotor, Prof. dr. ir. T. Houtgast, en mijn copromotor, Dr. ir. J. M. Festen voor de belangrijke bijdrage die zij aan dit proefschrift hebben geleverd. Beste Tammo, jij kwam altijd met creatieve oplossingen voor grote en kleine problemen waar ik tijdens het onderzoek tegenaan liep. Beste Joost, jij hebt met veel humor en energie de dagelijkse werkzaamheden in onze onderzoeksgroep begeleid. Jullie enthousiasme voor de wetenschap is zeer inspirerend.

Dit onderzoek was niet mogelijk geweest zonder de bereidwillige medewerking van de luisteraars die deelnamen aan mijn experimenten. Hartelijk dank voor het urenlang geduldig luisteren naar piepjes, woorden en zinnen.

De afdeling Audiologie van het VU-ziekenhuis is een stimulerende werkomgeving waar ik de afgelopen jaren met veel plezier heb gewerkt. Hierbij bedank ik alle medewerkers en vooral mijn kamergenoten, eerst Niek en toen Finn, voor de hulp en gezelligheid. Mijn 'klinische' collega's hebben me kennis laten maken met de dagelijkse gang van zaken in de kliniek en dat was een waardevolle ervaring. Mijn 'experimentele' collega's zijn een belangrijk wetenschappelijk klankbord geweest. Een paar van hen wil ik graag met name noemen. Niek, bedankt voor het wegwijs maken in de psychoakoestiek. Hans, dankjewel voor het geduldig beantwoorden van al mijn

computervragen en de hulp en tips bij het programmeren. Ingrid, we zijn ongeveer tegelijkertijd begonnen met onze promotie en hebben de verschillende stadia daarvan min of meer samen doorgemaakt. Hoewel onze projecten in beginsel verschillend waren, zijn de onderzoeken steeds meer naar elkaar toe gegroeid. Aan de discussies over wetenschap in het algemeen en de problemen van slechthorenden met het verstaan van spraak in het bijzonder heb ik veel gehad.

Als laatste bedank ik Guido en mijn ouders voor alle steun en betrokkenheid. Een promotie houd je tegen het einde toch wel heel erg bezig en ik ben blij voor alle begrip daarvoor. Mam, pap, heel erg bedankt dat jullie mij altijd gestimuleerd hebben om een antwoord te zoeken op al mijn vragen.

List of publications

van Schijndel, N. H, Thijssen, J. M., Oostendorp, T. F., Cuypers, M. H. M., and Huiskamp, G. J. M. (**1997**). "The inverse problem in electroretinography: A study based on skin potentials and a realistic geometry model," IEEE Transactions on Biomedical Engineering **44**(2), 209–211.

van Schijndel, N. H., Houtgast, T., and Festen, J. M., (**1998**). "Perceptual consequences of amplitude perturbations in the wavelet coding of speech," Proceedings of ICA/ASA 1998, Seattle, WA, 2599–2600.

van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**1999**). "Modeling intensity discrimination and detection in noise for stimuli with different spectro-temporal shapes," in *Psychophysics, Physiology and Models of Hearing*, edited by T. Dau, V. Hohmann, and B. Kollmeier (World Scientific Publishing, Singapore), pp. 165–168.

van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**1999**). "Intensity discrimination of Gaussian-windowed tones: Indications for the shape of the auditory frequency-time window," J. Acoust. Soc. Am. **105**, 3425–3435.

van Schijndel, N. H., Houtgast, T., and Festen, J. M., (**1999**). "Perceptual consequences of spectro-temporal smearing in the wavelet coding of speech," Proceedings of ASA/EAA/DEGA 1999, Berlin, 4APPB_2.

van Schijndel, N. H., Houtgast, T., and Festen, J. M., (**submitted**). "The effect of intensity perturbations on speech intelligibility for normal-hearing and hearing-impaired listeners," submitted to J. Acoust. Soc. Am..

van Schijndel, N. H., Houtgast, T., and Festen, J. M., (**submitted**). "Effects of degradation of intensity, time, or frequency content on speech intelligibility for normal-hearing and hearing-impaired listeners," submitted to J. Acoust. Soc. Am..

Curriculum vitae

- 16 april 1973: geboren te Heeswijk
- 1985–1991: Voorbereidend Wetenschappelijk Onderwijs aan Gymnasium Bernrode te Heeswijk
- 1991–1995: studie Natuurkunde aan de Katholieke Universiteit Nijmegen (cum laude). Afstudeerwerk werd verricht op de afdeling Medische Fysica en Biofysica onder begeleiding van Dr. T. F. Oostendorp en Dr.ir. J. Thijssen. Titel afstudeerscriptie: "Modeling the electroretinogram using realistic geometry."
- 1995–1999: Onderzoeker in Opleiding bij de afdeling Audiologie van het Academisch Ziekenhuis van de Vrije Universiteit te Amsterdam. Het onderzoek werd verricht onder leiding van Prof.dr.ir. T. Houtgast en Dr.ir. J. M. Festen. De resultaten zijn in dit proefschrift beschreven.
- 1999–heden: werkzaam als onderzoeker bij het Philips Natuurkundig Laboratorium te Eindhoven.