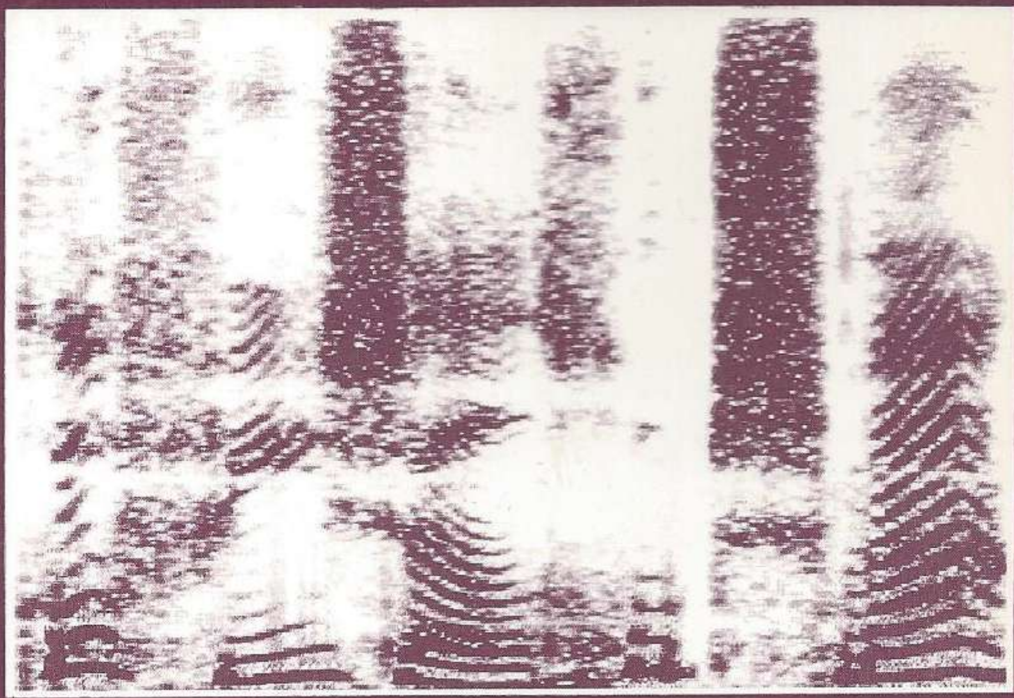
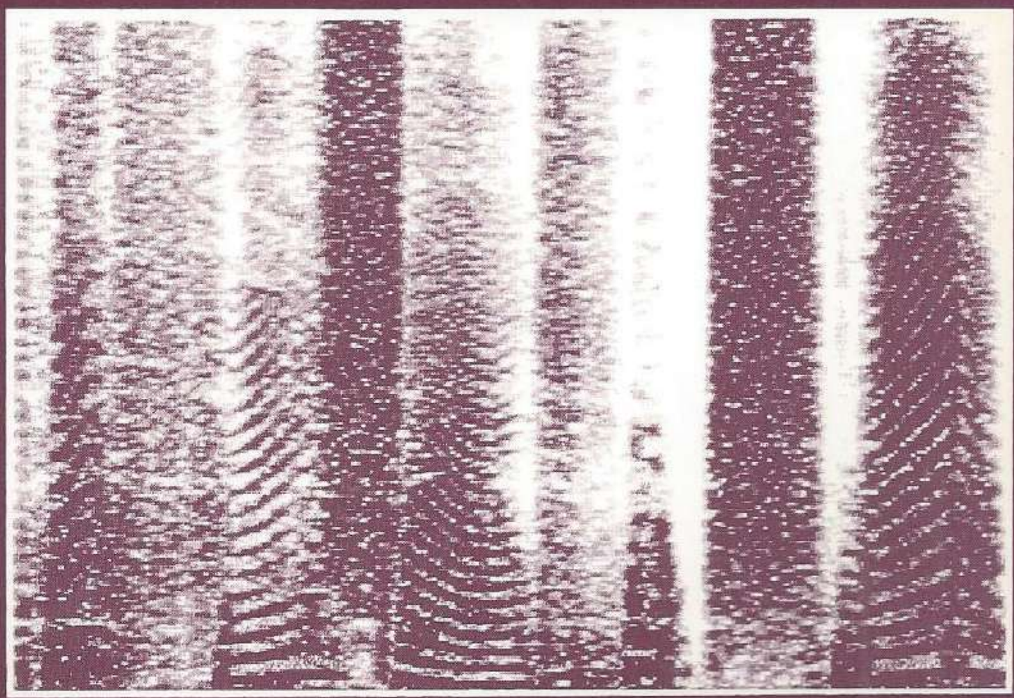
INTELLIGIBILITY OF SPECTRALLY SMEARED SPEECH





Mariken ter Keurs

INTELLIGIBILITY OF SPECTRALLY SMEARED SPEECH

VRIJE UNIVERSITEIT

INTELLIGIBILITY OF SPECTRALLY SMEARED SPEECH

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan
de Vrije Universiteit te Amsterdam,
op gezag van de rector magnificus
dr. C. Datema,
hoogleraar aan de faculteit der letteren,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der geneeskunde
op vrijdag 27 november 1992 te 10.30 uur
in het hoofdgebouw van de universiteit, De Boelelaan 1105

door

MARIKEN TER KEURS

geboren te Rijswijk (ZH)

Bariet, Ruinen
1992

Promotor: prof.dr.ir. R. Plomp
Copromotor: dr.ir. J.M. Festen
Referent: dr.ir. T. Houtgast

aan mijn ouders

VOORWOORD

Het in dit proefschrift beschreven onderzoek werd uitgevoerd bij de afdeling Experimentele Audiologie van de vakgroep Keel-, Neus-, en Oorheelkunde van de Vrije Universiteit te Amsterdam.

Graag wil ik iedereen die op enigerlei wijze heeft bijgedragen aan de totstandkoming van dit proefschrift, danken.

De volgende personen ben ik bijzonder erkentelijk en wil ik met name noemen. Prof. dr. ir. Reinier Plomp, mijn promotor, toonde steeds een warme belangstelling en begeleidde onderzoek en rapportage op de hem kenmerkende inspirerende en kritische wijze. Mijn copromotor, dr. ir. Joost Festen, leverde een aanzienlijke bijdrage aan de verwezenlijking van het onderzoek. Ing. Hans van Beek zorgde met grote inzet voor de zo belangrijke ondersteuning op soft- en hardware-matig gebied. Verder dank ik hen en alle andere collega's voor de plezierige werksfeer.

Tenslotte wil ik Sjaak, mijn familie en vrienden danken voor hun belangstelling en geduld.

CONTENTS

Chapter 1

GENERAL INTRODUCTION

The perception of speech is a complex process which, to a certain extent, depends on the peripheral auditory system's ability to resolve frequency. This suggests that there is a direct relation between auditory frequency resolution and speech intelligibility. As connected speech presented in quiet is highly redundant, the ear's high selectivity in frequency is thought to be of particular importance to understanding speech in a critical listening situation, for example, with interfering noise or competing speech.

While the ear's frequency resolving power itself has been studied quite extensively over the past two decades (for an overview, cf. Moore, 1986), and its relation to speech intelligibility is often stated, the degree of frequency resolution needed in speech discrimination has never been investigated quantitatively. A practical problem in establishing a direct relation between frequency selectivity and speech intelligibility is that the degree to which the ear is able to resolve frequency cannot be manipulated directly.

Generally, peripheral auditory frequency resolution is modelled as a bank of overlapping bandpass filters. The output of an auditory filter at a given frequency is determined by the shape of that filter applied to the power spectrum of an incoming sound. Consequently, reduced frequency resolution is modelled as a broadening of the auditory filters.

Reduced frequency resolution is thought to affect speech intelligibility roughly in two ways. Firstly, broadened auditory filters will preserve less spectral detail of the speech signal, which makes it more difficult to distinguish between the many spectrally based speech contrasts. With respect to spectral features, probably the most salient effect is a blurring of the spectral envelope. Secondly, in case of speech reception in the presence of interfering sounds, broadened filters will pass more of the background sound, which makes it more difficult to detect the speech signal. Indeed, limited frequency resolution has often been associated with hearing loss for speech in noise (e.g., Festen and Plomp, 1983; Dreschler and Plomp, 1985; Stelmachowitz *et al.*, 1985).

A possible approach to assessing the extent to which speech intelligibility in a critical listening situation depends on auditory frequency resolution, which bypasses the problem that the latter cannot be manipulated directly, is to decrease the amount of spectral information in the *speech signal* presented to the ear, and to measure the effect of this reduction on intelligibility. In this approach, those aspects of the speech signal are manipulated that, from the point of view of auditory frequency resolution are assumed to be important to the intelligibility of speech.

The aim of the present study is to quantify the importance of spectral contrasts to the intelligibility of speech in a critical listening situation. Apart from the purely scientific purpose of getting a better understanding of the spectral properties of speech and the involvement of auditory frequency resolution in speech reception, quantification of the importance of spectral contrast may also serve a more practical purpose, in that the results can be applied to, for example, the coding and transmission of speech signals, and to the signal processing performed to enhance the intelligibility of speech for hearing-impaired listeners.

The signal processing applied in the present study to achieve a reduction in spectral contrasts is based on an analysis-resynthesis system that consists of short-time fast Fourier transforms, spectral manipulation, and overlapping additions to reconstruct a continuous signal. Spectral contrasts are reduced in the speech signal by smearing the short-term spectral envelope by a convolution with a Gaussian-shaped filter on a log-frequency scale. In this way spectral contrasts can be gradually reduced, whereas the signal's phase and harmonic structure are preserved. The signal processing may be interpreted as the effect of limited resolution of the power-spectrum envelope isolated from auditory processing.

The masked speech-reception threshold (SRT) for sentences is chosen as the primary measure to assess the effect of reduced spectral contrasts on speech intelligibility. The masked SRT is defined as the speech-to-noise ratio in decibels at which 50% of short meaningful sentences, which are subjected to different degrees of spectral envelope smearing, can be reproduced by listeners without a single error (Plomp and Mimpen, 1979). The SRT is an accurate quantification of the intelligibility of the (spectrally smeared) speech in a critical listening situation, and because of its fixed performance criterion may be compared across experiments.

In the first two studies, chapters 2 and 3, a series of listening experiments is conducted with, in total, forty normal-hearing listeners to assess the general relation between spectral contrasts and speech intelligibility in a critical listenenig situation.

In these experiments the degree of spectral envelope smearing is systematically varied and the (noise or speech) masker is subjected to the same degree of spectral envelope smearing as the target speech. In chapter 2, the effect of spectral envelope smearing on the masked SRT for sentences pronounced by a female speaker is investigated. Additionally, the susceptibility of individual vowels and consonants to a reduction in spectral contrasts is assessed. In chapter 3, a possible differential effect of spectral envelope smearing on the intelligibility of male versus female speech is examined. Subsequently, the contribution of spectral contrasts to the rather large release from masking found for a speech masker relative to a noise masker is determined for both male and female target speech.

The results from chapters 2 and 3 point out a clear relation between spectral contrasts and speech intelligibility. An interesting question which follows, is whether limited resolution of spectral contrasts can explain (part of) hearing loss for speech in noise. This question is addressed in the final experiments described in this thesis, in chapter 4. Auditory filter bandwidth, as a measure of frequency resolution, is determined for fifteen sensorineurally hearing-impaired and eight normal-hearing listeners by estimating auditory filter shapes at center frequencies of 0.8, 1.6, and 3.2 kHz, using a notched-noise masking paradigm. The spectral contrasts essential to speech discrimination are determined by reducing the spectral contrasts in speech and measuring the reduction beyond which the SRT for sentences in noise increases. Correlation coefficients are computed between the various findings.

Chapter 5 summerizes the contents of this thesis. Chapters 2 to 4 are based on papers by ter Keurs, Festen, and Plomp (1992a, 1992b, 1992c) published, accepted, and submitted for publication in the Journal of the Acoustical Society of America, respectively.

EFFECT OF SPECTRAL ENVELOPE SMEARING ON SPEECH RECEPTION. I.*

ABSTRACT

The effect of reduced spectral contrast on the speech-reception threshold (SRT) for sentences in noise and on phoneme identification, was investigated with 16 normal-hearing subjects. Signal processing was performed by smoothing the envelope of the squared short-time fast Fourier transform (FFT) by convolving it with a Gaussian-shaped filter, and overlapping additions to reconstruct a continuous signal. Spectral energy in the frequency region from 100 to 8000 Hz was smeared over bandwidths of 1/8, 1/4, 1/3, 1/2, 1, 2, and 4 oct for the SRT experiment. Vowel and consonant identification was studied for smearing bandwidths of 1/8, 1/2, and 2 oct. Results showed the SRT in noise to increase as the spectral energy was smeared over bandwidths exceeding the ear's critical bandwidth. Vowel identification suffered more from this type of processing than consonant identification. Vowels were primarily confused with the back vowels /ɔ,u/, and consonants were confused where place of articulation is concerned.

*Paper published in: Journal of the Acoustical Society of America 91, 2872-2880 (1992a)

INTRODUCTION

Connected speech presented in quiet is highly redundant. In this condition, the resolution of spectral contrasts by the ear seems to be much larger than would be required. This suggests that the ear's high selectivity in frequency is of particular importance to understanding speech in the presence of interfering sounds. Indeed, a significant correlation has been found between speech perception in noise and measures of frequency resolution (Rosen and Fourcin, 1986).

The ear's resolving power has been measured quite extensively over the past two decades and its relation to speech intelligibility is often stated (Moore, 1986). However, the relation itself, that is the relation between reduced frequency selectivity and the resolution of spectral contrasts needed in speech discrimination, has not been clearly established.

A major problem in establishing a direct relation between frequency selectivity and speech intelligibility is that the ear's ability to resolve frequency cannot be manipulated directly. Sensorineurally hearing-impaired subjects have been used to elucidate the problem but this approach is rather risky, as sensorineural hearing impairment seldom consists of a loss of frequency resolution only. Moreover, a difficult problem when comparing normal and hearing-impaired listeners is the presentation level of the stimuli, especially because many psychoacoustical phenomena depend to some degree on stimulus level (e.g., Dubno and Schaefer, 1989; Tyler, 1986).

Quite a different approach, is an attempt to isolate frequency resolution from other factors that influence speech intelligibility by processing the *speech signal*. In this approach one assumes a reduction of spectral contrasts in the speech signal to cause a loss in perception by normal-hearing listeners in the same degree as though their auditory bandwidths were actually broadened.

A limited number of studies have used such an approach. However, in most cases, the investigation of the effects of spectral blurring on speech perception was a subordinate part of an attempt to model more generally aspects of sensorineural hearing impairment. Villchur (1977), for example, blurred speech-frequency elements by coding the speech with a 16-channel amplitude-modulated noise-carrier system. The center frequencies and widths of the noise bands were adjustable, and the noise bands were allowed to overlap one another in frequency. When the bandwidth of each of the 16 noise bands was increased to 2⅓ oct, sentence intelligibility was judged by the experimenter to be badly impaired but not entirely destroyed. More recently, Gagné and Erber (1987) simulated reduced frequency selectivity and discrimination by passing the signal through a time/frequency jittering device that effectively increased the bandwidth of the signal. The quality of the speech suffered from the processing, but a clear effect on intelligibility could not be established. Howard-Jones *et al.* (1991) used three different types of analog processing and a digital one to degrade the frequency content of speech. Again, spectral blurring seemed to have a detrimental effect on speech discrimination.

A more quantitative approach was used by Celmer and Bienvenue (1987). They modified the short-term speech spectrum by dividing it in fixed consecutive frequency regions, and setting the amplitude of the discrete spectral components in each region to the regional root-mean square. With increasing processing bandwidth masked word intelligibility scores decreased.

Generally, speech discrimination (in noise) seems to decrease with deteriorating spectral contrasts. It is obvious that more subtle approaches are needed to prove a direct relation between reduced frequency selectivity and the resolution of spectral contrasts needed in speech reception.

In the present paper, we investigated in listening experiments the extent to which speech intelligibility in noise depends on spectral contrasts in the signal. Loss of frequency resolution of the auditory system was simulated by smearing the short-term spectral envelope of the speech signal on a log-frequency scale with a Gaussian-shaped filter. In this way spectral contrasts were gradually reduced, but the signal's phase and harmonic structure were preserved. In the first experiment the speech-reception threshold (SRT) for sentences in noise was measured as a function of spectral envelope smearing. Effects on vowel and consonant identification were studied in a second and third experiment.

I. METHOD

A. Signal processing to reduce spectral contrasts : Spectral envelope smearing

The signal processing is based on an analysis-resynthesis scheme that allows for more or less arbitrary modification of the short-term spectrum in either its phase or amplitude content without introducing undesired distortion in the resynthesized signal. It consists of short-time fast Fourier transforms (FFT) and overlapping

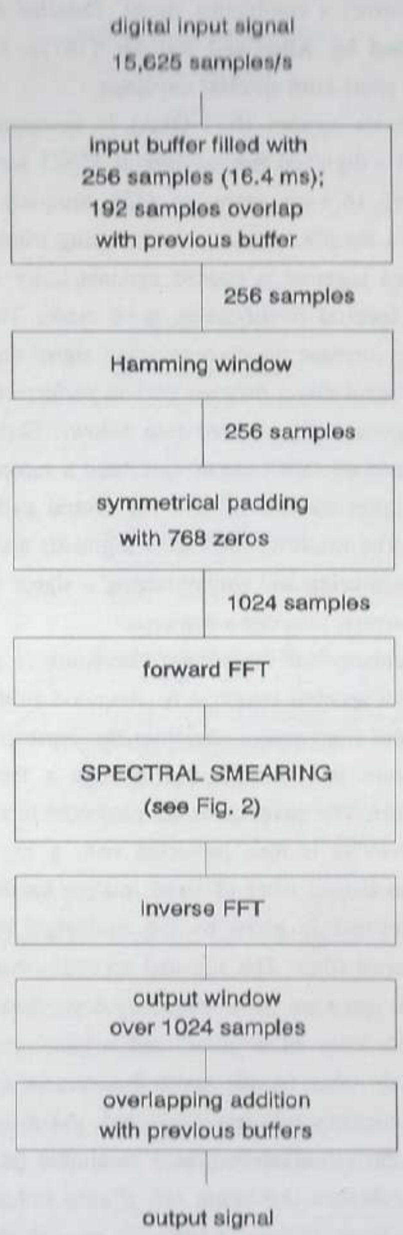


FIG. 1. Schematic representation of the analysis-resynthesis system.

additions (OLA) to reconstruct a continuous signal. Detailed conditions for such a system have been described by Allen and Rabiner (1977). Spectral contrasts are reduced by smoothing the short-term spectral envelope.

The analysis-resynthesis system (FFT-OLA) is represented schematically in Fig. 1. The speech signal - digitized into 16 bits at 15625 samples per second - is sectioned into overlapping 16.4-ms segments (256 samples). The shift between consecutive segments is 4.1 ms (64 samples). A Hamming window is applied to each segment and the windowed segment is padded symmetrically with 768 zeros (49.1 ms) to accommodate the spectral modification to be made. This is necessary since spectral modifications may increase the corresponding signal duration. On the whole segment of 1024 samples (65.6 ms) a forward FFT is performed and the envelope of the resulting short-term spectrum is smeared (see below). Each processed spectrum is then inversely transformed into the time domain, and a second window is applied. This window of 1024 samples consists of an unity central part of 256 samples and sine wave-squared skirts. The resulting successive segments are added to reconstruct a continuous signal. Just analyzing and resynthesizing a signal with this system does not, apart from rounding errors, alter the signal.

A schematic representation of the spectral smearing is given in Fig. 2. For each segment the short-term spectral envelope is computed on the basis of the power spectrum. From the discrete (log) power spectrum the cepstrum is calculated by an inverse FFT. This cepstrum is "liftered" and through a forward FFT the (log) spectral envelope is obtained. The envelope is then adjusted to run smoothly over the spectral peaks. This envelope is then projected onto a log-frequency scale and convolved with a Gaussian-shaped filter of fixed relative bandwidth. The degree to which the envelope is smeared is given by the equivalent rectangular bandwidth (ERB) of the Gaussian-shaped filter. The smeared spectral envelope is then imposed upon the original complex spectrum by a frequency-dependent multiplication of the original spectrum with the ratio of smeared and original envelope. In this way spectral energy is smeared - that is, the spectral contrasts are reduced - but the signal's phase and harmonic structure are preserved. As it is unclear how phase would be affected by the ear's limited frequency resolution (cf. Moore, 1986), and there is no experimental evidence that phase (cf. Plomp and Steeneken, 1969) and resolution of harmonics have noticeable effects on speech intelligibility, this smearing process is, as far as speech intelligibility is concerned, assumed to be a reasonable simulation of the ear's limited frequency resolution.

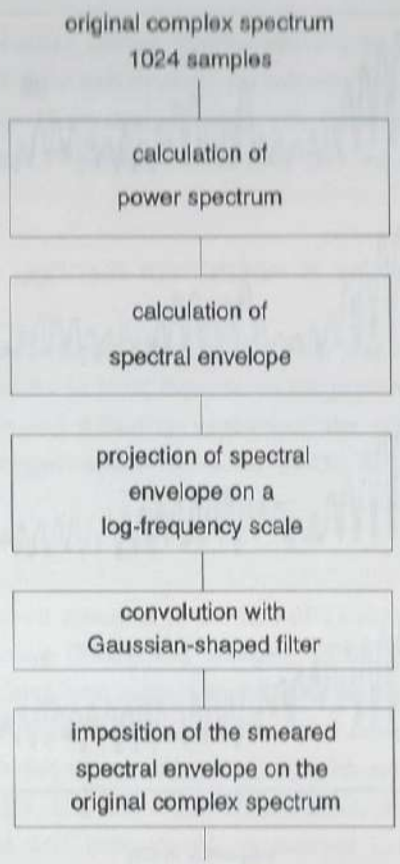


FIG. 2. Schematic representation of the spectral envelope smearing procedure.

Figure 3 gives an example of the effect of spectral envelope smearing on an arbitrary speech segment in 3 conditions. From top to bottom the figure shows the original log-power spectrum and the log-power spectrum of the segment smeared over an ERB of 1/8, 1/2, and 2 oct, respectively. The smearing causes rapid fluctuations in the spectral envelope to disappear first. When the smearing bandwidth increases, ever slower fluctuations follow. As the smearing is performed on the spectral envelope it works just as well for noisy as for periodic signals.

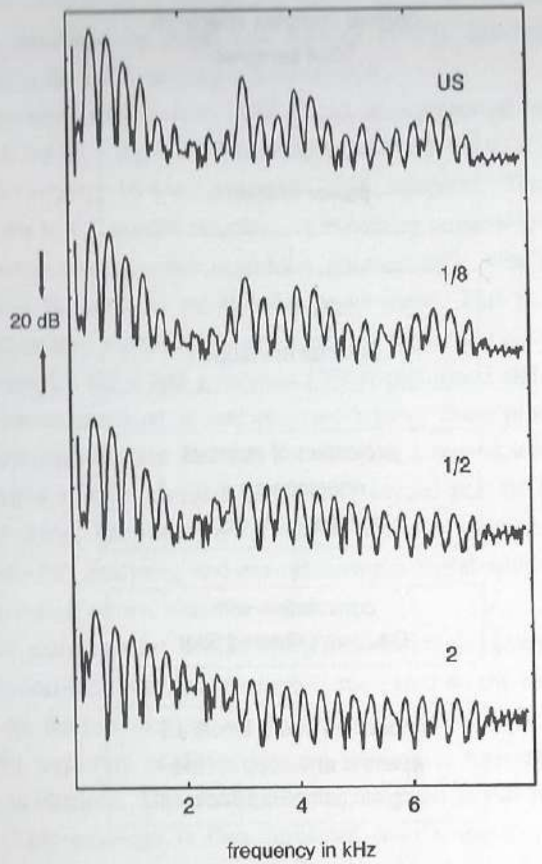


FIG. 3. Log-power spectra of an arbitrary speech segment as a function of spectral envelope smearing. From top to bottom: the original log spectrum and the log spectra after smearing over 1/8, 1/2, and 2 oct, respectively.

B. Materials

The stimuli used in the three experiments - sentences, monosyllables and vowels in isolation - were digitized into 16 bits at a sample frequency of 15625 Hz with four times oversampling, and digitally low-pass filtered at 6250 Hz. Filter-rejection slopes exceeded 80 dB/oct. All stimuli were pronounced in a normal fashion by a trained female voice.

C. Subjects

Sixteen normal-hearing college students between the ages of 18 and 24 served as paid listeners in all three experiments. All subjects had a pure-tone hearing level in both ears lower than 15 dB over the range 250 Hz to 8 kHz, and were free of a history of ear pathology. The order of the experiments was varied over listeners.

II. EXPERIMENT 1 : SPEECH RECEPTION IN NOISE

As stated earlier the primary goal of this study was to investigate the extent to which speech intelligibility in noise depends on the presence of spectral contrasts in the speech signal. In the following experiment the effect of spectral envelope smearing on the speech-reception threshold (SRT) for sentences in noise was investigated.

A. Stimuli

The speech material consisted of ten lists of 13 short (eight or nine syllables) everyday Dutch sentences (Plomp and Mimpen, 1979). A masking noise with the same spectrum as the long-term average spectrum of the sentences was used.

Eight conditions were investigated : the signal analyzed and resynthesized by the FFT-OLA system but unsmeared (US), and with seven conditions of spectral smearing with ERBs 1/8, 1/4, 1/3, 1/2, 1, 2, and 4 oct, respectively.

The (long-term) S/N ratio should be constant over the spectrum in each condition of spectral smearing. Otherwise, depending on the spectral smearing, some frequency regions in the spectrum may be masked more than others. Therefore, to preclude any unwanted masking both speech and noise were subjected to the same degree of spectral smearing in each condition. Analysis of the long-term average spectra of the sentences and the noise after processing showed them to be similar. The processing used to accomplish the reduction in spectral contrast includes a nonlinear component: The estimation of the short-term spectral envelope. Therefore, speech and noise must be processed separately to prevent the smearing depending on the S/N ratio of speech mixed with noise.

B. Test procedure

The subjects listened to the lists of sentences monaurally (right ear) through headphone (Beyer DT 48) in a soundproof room. The lists were presented in a fixed order. The sequence of presentation of the eight conditions was varied over listeners and lists according to a 8x8 digram-balanced Latin square to minimize order and list effects. As there were eight sequences and 16 subjects, each sequence was heard by two subjects. Prior to the administration of the test lists, the subjects were presented with a sentence list spectrally smeared over 1 oct and a list smeared over 4 oct to familiarize them with the quality of the speech and the experimental task.

The level of the masking noise was fixed at 65 dBA and the level of the sentences was changed according to an up-down adaptive procedure (Plomp and Mimpen, 1979). The masked SRT was defined as the speech-to-noise ratio in decibels at which 50% of the sentences could be reproduced without a single error. In the adaptive procedure, the first sentence of each list was presented at a level below the reception threshold. This sentence was then presented repeatedly, each time at a 4-dB-higher level, until the listener was able to reproduce it correctly. The remaining 12 sentences in a list were then presented only once in a simple up-down procedure with a step size of 2 dB. An errorless reproduction of the entire sentence was required for a correct response. The average presentation level over sentences 4-13 was taken as the SRT.

C. Results and discussion

Figure 4 shows the mean SRT for sentences in noise as a function of spectral envelope smearing. The SRTs for smearing conditions of 1/8, 1/4, and 1/3 oct are comparable with the SRT for unsmeared speech. This shows that for these conditions the processing does not affect intelligibility, while in fact there is already some loss of spectral contrast. The average threshold of about -4.5 dB agrees with the threshold found for normal speech in earlier studies (Plomp, 1986).

From a bandwidth of 1/3 oct upwards the SRT increases progressively. The elevation in threshold between the 1/3- and 1/2-oct condition is just significant (two-tailed t test, $p<0.04$); from a 1/2 oct upwards all elevations between successive smearing bandwidths are highly significant (two-tailed t test, $p<0.004$).

The standard deviation of the interindividual differences in SRT for conditions up to 2 oct is on the average 1.4 dB. This is only slightly higher than the standard deviation of about 1 dB of the measuring procedure (cf. Plomp and Mimpen, 1979). This shows that for these conditions the subjects are quite alike in their ability to understand speech with reduced spectral contrast. However, in the 4-oct condition a

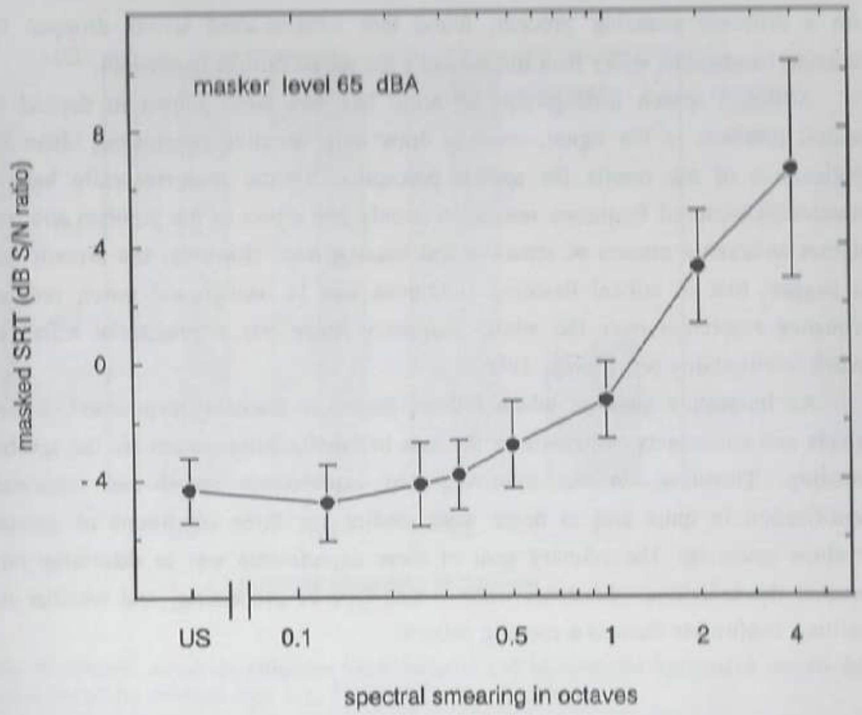


FIG. 4. Mean speech-reception threshold for spectrally smeared sentences in noise, expressed in signal-to-noise ratio, as a function of smearing bandwidth. Vertical bars represent the standard deviation of the SRT in each condition.

clearly increased interindividual spread of 3.7 dB is found. Apparently, when the spectrum is highly smeared, other factors such as the ability to make use of temporal cues determine whether the sentences are understood.

These data confirm the threshold shifts found in a pilot study on the subject (ter Keurs *et al.*, 1989a, 1989b). Similar conditions were tested with a slightly different smearing process. Elevated SRTs were found for smearing bandwidths of a 1/2 oct and upward, and the transition between the threshold plateau and the elevated thresholds was somewhere between 1/4 and 1/2 oct (the 1/3-oct condition was not included). In the present study, the transition lies between 1/3 and 1/2 oct. This bandwidth approximates the ear's critical bandwidth which lies between 1/4 and 1/3 oct. A similar inflection point was found by Celmer and Bienvenue (1987) who, with a different smearing process, found that masked-word scores dropped for smearing bandwidths wider than the subject's measured critical bandwidth.

Although speech intelligibility in noise has now been shown to depend on spectral contrasts in the signal, one can draw only tentative conclusions about the implications of the results for speech perception by the sensorineurally hearing impaired, as reduced frequency resolution is only one aspect of the problem and may interact with other aspects of sensorineural hearing loss. However, the present data do suggest that in critical listening conditions like in background noise, reduced frequency resolution over the whole frequency range has a substantial effect on speech intelligibility (cf. Plomp, 1986).

An interesting question which follows from the preceding experiment is how vowels and consonants contribute to the loss in intelligibility caused by the spectral smearing. Therefore, in the following two experiments vowel and consonant identification in quiet and in noise were studied for three conditions of spectral envelope smearing. The primary goal of these experiments was to determine how resistant the individual phonemes were to this type of processing, and whether the resulting confusions showed a specific pattern.

III. EXPERIMENT 2 : VOWEL IDENTIFICATION

A. Stimuli

The stimuli consisted of 8 utterances of 11 Dutch monophthongs /a,ɑ,e,ɛ,i,I,o,ɔ,u,y,œ/. The vowels were equalized in SPL and length. Total length of the stimuli including rise and fall times of 12.5 ms was 115 ms.

B. Test procedure

Unsmeared vowels, and vowels processed in three conditions of spectral envelope smearing - 1/8, 1/2, and 2 oct - were presented for identification both in quiet and in noise. Conditions were selected on the basis of a pilot SRT experiment (ter Keurs *et al.*, 1989a, 1989b), and were chosen to reflect the normal intelligibility plateau (unsmeared and 1/8 oct), the onset of reduced intelligibility (1/2 oct), and low intelligibility (2 oct), respectively. In the masked conditions, noise with the long-term average spectrum of the sentences of experiment 1 was used. All vowel stimuli were presented binaurally through headphone (Beyer DT 48) at a level of 65

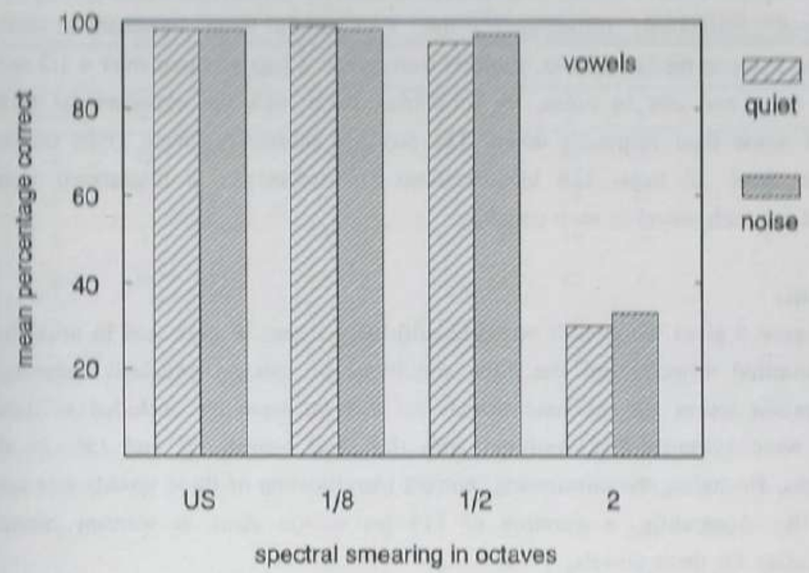


FIG. 5. Overall vowel-identification score in quiet and in noise for unsmeared vowels and vowels spectrally smeared over 1/8, 1/2, and 2 oct, respectively.

dBA. The level of the noise was fixed at 60 dBA; that is, the masked vowel stimuli were presented at a S/N-ratio of +5 dBA. At this S/N ratio, 10 dB above the SRT for unsmeared sentences, a maximal identification score of about 90% was expected for unprocessed vowels (cf. Bosman, 1989, Chap. 5). In each condition, both vowels and noise were subjected to the same degree of spectral smearing, and vowels and noise were processed separately.

From the vowel material (eight repetitions of the eight utterances x 11 vowels), 32 semirandom lists of 22 vowels were constructed containing two utterances of each vowel. The test stimuli (32 lists x 4 smearing conditions) were presented to 16 subjects over two blocks in quiet and two blocks in noise. In each block, the four smearing conditions were tested over eight lists (two consecutive lists per condition). The lists within each block were presented in a fixed order. The sequence of the four smearing conditions in each block varied over listeners and lists according to a 4x4 digram-balanced Latin square. The order of the blocks (i.e., whether the subjects started with the masked or quiet condition, or with the first or second block in quiet or noise) was counterbalanced. In the masked condition, noise was given continuously throughout a list. The stimuli were presented at 2.5-s intervals. Prior to the test blocks, subjects were given 2 lists smeared over a 1/2 oct, one in quiet and one in noise, to familiarize them with the experimental task. Subjects wrote their responses down. The possible alternatives were given on the response sheet. In total, 128 identifications (16 subjects x 8 utterances) were obtained for each vowel in each condition.

C. Results

Figure 5 gives the overall vowel-identification scores in quiet and in noise for the unsmeared vowels, and the three conditions of spectral envelope smearing. Identification scores for the long vowels /e/ and /o/ were not included as these vowels were systematically confused with the short vowels /I/ and /ɔ/. In all conditions, (including the unsmeared), correct identification of these vowels was less than 30%. Apparently, a duration of 115 ms is too short to warrant correct identification for these vowels.

Overall identification for unsmeared vowels and for vowels smeared over 1/8 and 1/2 oct was 97%, both in quiet and in noise. Identification dropped dramatically to 32% in the 2-oct condition. Unfortunately, the ceiling effect in the first three conditions makes an estimation of the critical smearing bandwidth impossible.

Confusion matrices for the 2-oct condition in quiet and in noise are given in Table I. Spectral envelope smearing of the 11 vowels - including back, middle and front vowels - resulted in a predominant proportion of vowels being identified as the back vowels /ɔ/ and /u/. The patterns of confusions in the quiet and masked condition were very similar. Therefore, in the following discussion data are taken from the quiet condition, but the results apply equally well to the masked condition.

Generally, responses corresponded to vowels with low first and second formants ($F1$ and $F2$), and within this subset, vowels were chosen whose $F1$ was closest in frequency to the F1 of the target vowel [formant frequencies of Dutch vowels (female) as given by van Nierop *et al.*, 1973]. In particular, 73% (90/124) of occurrences of the target vowel /i/ and 78% (100/128) of occurrences of the target vowel /y/ were identified as /u/, while target vowel /u/ itself was correctly identified 95% (121/128) of the time. Target vowels with higher $F1$, /a,ɑ,e,ɛ,I,o,œ/, were most often identified as /ɔ/. The frequency of $F1$ of these target vowels seemed to determine the degree to which they were confused with /ɔ/. For example,

TABLE I. Confusion matrices for vowels smeared over 2 oct. Scores for the quiet condition are given in the top matrix and scores for the masked condition in the bottom one.

	responses						VOWELS IN QUIET					
stimuli	a	ɑ	e	ɛ	i	I	o	ɔ	u	y	œ	sum
a	40	36	.	3	.	.	7	44	2	.	.	132*
ɑ	9	44	.	.	.	.	7	58	10	.	.	128
e	3	.	.	4	.	4	16	89	10	1	1	128
ɛ	10	41	.	5	.	.	5	62	4	.	1	128
i	.	.	.	.	9	1	4	16	90	3	1	124*
I	.	2	.	3	.	8	8	79	19	6	3	128
o	1	3	.	.	.	.	22	86	13	1	2	128
ɔ	.	1	.	.	.	.	11	98	16	2	.	128
u	.	.	.	.	.	.	3	3	121	1	.	128
y	.	.	.	.	5	.	2	11	100	10	.	128
œ	.	.	.	3	1	2	17	65	24	10	6	128
sum	63	127	.	18	15	15	102	611	409	34	14	1408

	responses						VOWELS IN NOISE					
stimuli	a	ɑ	e	ɛ	i	I	o	ɔ	u	y	œ	sum
a	23	39	.	7	.	5	7	40	6	1	4	132
ɑ	6	38	.	3	.	1	4	62	9	2	3	128
e	.	2	2	9	1	15	8	67	15	4	5	128
ɛ	2	15	.	13	.	6	3	80	5	1	3	128
i	.	.	1	.	26	7	.	6	74	8	2	124
I	.	1	.	5	6	19	6	52	31	4	4	128
o	.	1	.	.	.	.	18	70	34	1	4	128
ɔ	.	3	.	.	.	1	4	89	29	2	.	128
u	.	.	.	.	.	.	.	3	125	.	.	128
y	.	.	.	.	14	2	.	5	74	28	5	128
œ	.	.	.	1	2	5	6	51	35	12	16	128
sum	31	99	3	38	49	61	56	525	437	63	46	1408

*Through an oversight, 132 and 124 identifications were obtained for /a/ and /i/, respectively.

identification of target vowel /a/, the vowel with the highest $F1$, ran from /a/ (30%, 40/132), via /a/ (27%, 36/132) to /ɔ/ (33%, 44/132). At the other end, 70% (89/128) of occurrences of target vowel /e/, whose $F1$ is comparable with that of /ɔ/, were confused with /ɔ/. Smearing the spectral envelope of the vowels results in an envelope of which the high-frequency ripples are increasingly smoothed out. In the 2-oct condition the ripples characterizing $F2$ and higher formants have been effectively smoothed out, leaving a broadened $F1$ to primarily determine identification. It is plausible that increasing the smearing bandwidth to more than 2 oct will eventually result in all vowels being identified as the back vowel /u/.

The results accord with the findings of Delattre *et al.* (1952), and Dubno and Dorman (1987). Delattre *et al.* determined vowel identification for two-formant vowels of which the second formant was reduced in intensity. This resulted in the target vowels being identified as those particular back vowels whose first formant the target vowels shared. Dubno and Dorman determined front-vowel identification using synthetic stimuli whose first formants were specified in a normal manner and whose higher formants were either absent or represented by a flat low-amplitude region of energy in the vicinity of the second formant. A large proportion of the stimuli with only one formant was identified as back vowels. Consistently, the incorrect identification was a vowel whose $F1$ was closest in frequency to the first formant of the target vowel. The stimuli with a broad low-amplitude region of energy in the vicinity of the second formant yielded very good identification scores. This is consistent with the identification score we obtained for vowels spectrally smeared over a 1/2 oct. Likewise, in this condition there is energy left in the region of the $F2$, and apparently, at the presentation level used this is enough to warrant normal identification.

IV. EXPERIMENT 3 : CONSONANT IDENTIFICATION

A. Stimuli

The speech material consisted of 16 lists of 12 Dutch meaningless syllables of the consonant-vowel-consonant type (CVC syllables) as designed by Bosman (1989, Chap. 3). Each list consisted of 12 different syllables and each syllable was composed of an initial consonant, a vowel, and a final consonant chosen from three sets of 12 phonemes. All phonemes of a set appeared only once in each list. The set

of initial consonants consisted of /t,k,b,d,X,v,z,n,l,j,w,h/, the set of vowels of /a,ɑ,e,ɛ,i,I,o,ɔ,ϕ,u,au,ɛi/, and the set of final consonants of /p,t,k,f,s,X,m,ŋ,n,l,j,w/. A major advantage of this material is that the consonants appear with different vowels. The syllables were normalized for rms level.

B. Test procedure

Apparatus and stimulus-presentation paradigm were similar to those used in experiment 2. Unsmeared syllables and syllables processed in three conditions of spectral smearing - 1/8, 1/2, and 2 oct - were tested both in quiet and in noise. In order to avoid a bias of sense responses to nonsense stimuli, subjects were asked to identify only either initial or final consonant of the presented syllables. Once again, masking noise with the same spectrum as the long-term average spectrum of the sentences was used. In each condition, both speech and noise were subjected to the same degree of spectral smearing, and speech and noise were processed separately. All stimuli were presented binaurally at a level of 65 dBA through headphone (Beyer DT 48). In the masked condition the noise was fixed at 60 dBA.

From the original 16 lists of 12 syllables 32 semirandom lists of 24 syllables were constructed (each new list was composed of two original lists of which the sequence of the syllables was semirandomized). The test stimuli in four smearing conditions and two consonant conditions (initial versus final), were presented to 16 subjects over two blocks in quiet and two blocks in noise. Both in quiet and in noise, one block was scored for initial consonant and one for final. In each block, the four smearing conditions were tested over eight lists. The lists within each block were presented in a fixed order. The sequence of the four smearing conditions in each block varied over listeners and lists according to a 4x4 digram-balanced Latin square. The order of the blocks (i.e., whether the subjects started with the masked or quiet condition, or with initial or final consonant) was counterbalanced. In the masked condition, noise was presented continuously throughout a list. The stimuli were presented at 2-s intervals. Prior to the test blocks subjects were given two lists smeared over a 1/2 oct, one in quiet to be scored for initial consonant and one in noise to be scored for final consonant, to familiarize them with the experimental task. The subjects wrote their responses down, and were allowed an open-response set. In this way, 64 identifications (16 subjects x 4 repetitions of the syllables) were obtained for each initial and final consonant in each condition.

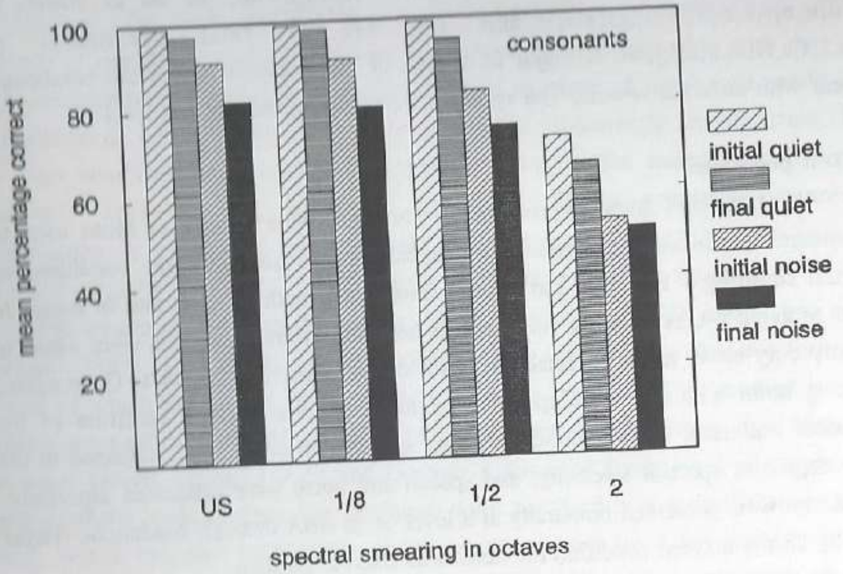


FIG. 6. Overall identification score for initial and final consonants in quiet and in noise for unsmeared consonants and consonants spectrally smeared over 1/8, 1/2, and 2 oct, respectively.

C. Results

Figure 6 gives the overall identification score for initial and final consonants in quiet and in noise for the unsmeared condition, and the three conditions of spectral envelope smearing. Overall identification in quiet for unsmeared initial consonants, and initial consonants smeared over 1/8 and 1/2 oct, was 99%. Identification dropped to 72% in the 2-oct condition. Overall identification of final consonants in quiet was 97% for the first three conditions, and the score dropped to 66% in the 2-oct condition. Scores for initial consonants in noise dropped from an average 91% in the first two conditions (unsmeared and 1/8 oct) to 83% in the 1/2-oct condition, and 53% in the 2-oct condition. Scores for masked final consonants dropped from an average 81% in the first two conditions to 75% and 51% in the last two conditions. The bandwidth beyond which the scores dropped lies somewhere between 1/8 and 1/2 oct. This agrees rather well with the inflection bandwidth found for the SRT in noise, which has a width between 1/3 and 1/2 oct.

The average identification score for consonants spectrally smeared over 2 oct was 61%. This is nearly twice the average vowel identification score (32%) in the same condition. Apparently, in this condition, consonant identification is much less susceptible to loss of spectral contrasts than is vowel identification. This is not surprising as consonant identification is known to depend only partially on spectral cues, and the harmonic structure of the signal (voiced-voiceless distinction) is preserved.

Confusion matrices for initial and final consonants in the 2-oct conditions in quiet and in noise are given in Table II. Spectral envelope smearing of the consonants consistently resulted in place-of-articulation errors. This applies to both initial and final consonants, and to the quiet and masked condition. It seems that primarily the spectral information cueing the articulatory point of constriction of the target stimuli is affected by the processing or by masking. Thus, plosives, fricatives, nasals, and semivowels were predominantly confused with consonants within their respective categories. A similar result was found by Miller and Nicely (1954), who analyzed the pattern of perceptual confusions resulting from low-pass-filtered consonants.

The degree to which each individual consonant suffered from spectral smearing depends both on the degree to which it is characterized spectrally and on its distribution of energy over the spectrum. For example, the fricatives /s/ or /z/ are spectrally characterized by energy at high frequencies, whereas the fricatives /f/ and /v/ are characterized by energy spread uniformly over the spectrum (broad-band spectrograms of consonants as given by Nooteboom and Cohen, 1984). As a result of spectral smearing the high-frequency energy of /s/ and /z/ is spread over the spectrum, while the energy distribution of /f/ and /v/ remains relatively unaffected. Consequently, /s/ and /z/ were increasingly identified as /f/ and /v/. Another example is the identification of the plosives /b/ and /d/; /b/ is spectrally characterized by energy at low frequencies and /d/ by energy at higher frequencies. As a result, identification of target stimulus /b/ remained relatively unaffected by spectral smearing, whereas target stimulus /d/ was increasingly identified as /b/.

TABLE II. Confusion matrices for consonants smeared over 2 oct. The matrices show scores for initial consonants in quiet and in noise, and final consonants in quiet and in noise, respectively.

INITIAL CONSONANTS IN QUIET

stimuli \ responses	t	k	p	b	d	v	f	z	s	X	n	m	l	w	j	h	r	sum
t	60	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	64
k	8	27	28	.	.	.	.	.	.	.	.	.	.	.	1	.	.	64
b	.	.	.	63	1	.	.	.	.	.	.	.	.	.	.	.	.	64
d	.	.	.	28	36	.	.	.	.	.	.	.	.	.	.	.	.	64
v	.	.	.	.	.	56	2	.	.	.	.	.	.	6	.	.	.	64
z	.	.	.	.	.	25	2	32	2	.	.	.	.	3	.	.	.	64
X	.	.	.	.	.	9	14	.	.	41	.	.	.	.	.	.	.	64
n	.	.	.	.	.	.	.	.	.	.	61	1	1	1	.	.	.	64
l	.	.	.	.	.	.	.	.	.	.	6	.	54	2	.	1	1	64
w	.	.	.	.	.	1	.	.	.	.	.	.	.	62	.	.	1	64
j	.	.	.	.	.	1	.	.	.	.	.	.	.	29	32	1	1	64
h	.	.	.	.	.	36	.	.	.	.	.	.	.	3	.	25	.	64
sum	68	27	32	91	37	128	18	32	2	41	67	1	55	106	33	27	3	768

INITIAL CONSONANTS IN NOISE

stimuli \ responses	t	k	p	b	d	v	f	z	s	X	n	m	l	w	j	h	r	sum
t	55	1	8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	64
k	5	21	32	2	.	2	1	.	.	1	.	.	.	.	.	.	.	64
b	.	2	.	43	5	.	.	.	.	.	.	.	.	10	.	3	1	64
d	1	.	3	22	23	2	1	.	.	.	.	.	.	5	1	6	.	64
v	.	1	1	1	.	38	1	2	.	1	.	.	.	12	3	2	2	64
z	.	.	.	.	.	27	1	33	.	.	.	.	.	3	.	.	.	64
X	.	.	.	.	.	20	7	.	1	36	.	.	.	.	.	.	.	64
n	.	.	.	.	.	.	.	.	.	.	52	4	5	2	.	.	1	64
l	.	.	.	.	.	.	.	.	.	.	17	1	31	8	3	2	2	64
w	.	.	.	21	2	3	1	.	.	.	.	.	.	33	.	2	2	64
j	.	1	.	3	1	4	.	.	.	.	4	1	5	26	16	1	2	64
h	.	.	.	.	.	21	2	2	.	.	.	.	.	12	.	25	2	64
sum	61	26	44	92	31	117	14	37	1	38	73	6	41	111	23	41	12	768

FINAL CONSONANTS IN QUIET

stimuli \ responses	t	k	p	b	d	f	s	X	η	n	m	l	j	w	sum
t	50	1	13	.	.	.	.	.	.	.	.	.	.	.	64
k	1	17	45	1	.	.	.	.	.	.	.	.	.	.	64
p	.	.	62	2	.	.	.	.	.	.	.	.	.	.	64
f	.	.	.	.	.	63	.	.	.	.	.	.	.	1	64
s	.	.	.	.	.	29	34	.	.	.	.	.	.	1	64
X	.	.	.	.	1	32	.	31	.	.	.	.	.	.	64
η	1	.	.	.	1	.	.	.	56	2	1	3	.	.	64
n	.	.	.	.	.	1	.	.	12	40	6	4	.	1	64
m	.	.	.	.	.	.	.	.	8	28	22	6	.	.	64
l	.	.	.	.	.	.	.	.	2	2	.	50	3	7	64
j	.	.	.	.	.	1	.	.	3	.	.	4	54	2	64
w	.	.	.	.	.	.	.	.	.	.	.	38	.	26	64
sum	52	18	120	3	2	126	34	31	81	72	29	105	57	38	768

FINAL CONSONANTS IN NOISE

stimuli \ responses	t	k	p	b	f	s	X	η	n	m	l	r	j	w	h	sum
t	56	2	4	.	2	.	.	.	.	.	.	.	.	.	.	64
k	6	16	36	.	2	.	2	2	.	.	.	.	.	.	.	64
p	5	8	45	1	1	.	.	.	1	1	1	.	.	1	.	64
f	1	1	3	.	50	2	2	.	1	.	1	2	.	1	.	64
s	.	.	.	.	17	46	.	.	.	.	.	.	1	.	.	64
X	2	.	1	.	30	3	26	.	.	.	.	2	.	.	.	64
η	.	.	2	.	3	.	.	26	10	6	6	.	5	6	.	64
n	.	.	1	.	.	.	1	10	26	12	6	.	1	7	.	64
m	.	.	1	.	.	.	.	8	34	13	4	1	2	1	.	64
l	.	.	2	.	1	.	.	1	12	1	30	1	10	6	.	64
j	.	.	1	.	1	.	.	3	5	1	11	.	37	5	.	64
w	.	.	1	.	1	.	.	.	8	4	23	2	4	20	1	64
sum	70	27	97	1	108	51	31	50	97	38	82	8	60	47	1	768

V. CONCLUSION

In the present paper, we investigated the extent to which speech intelligibility in noise depends on the resolution of spectral contrasts by the ear. Loss of frequency resolution of the auditory system was simulated by smearing the short-term spectral envelope of the speech signal on a log-frequency scale with a Gaussian-shaped filter. Results showed the SRT in noise to increase as the spectral energy was smeared over bandwidths exceeding the ear's critical bandwidth. As would be expected, vowel identification appears to depend more on spectral contrasts and their resolution than consonant identification. Vowels were primarily confused with the back vowels /ɔ,u/, and consonants were confused where place of articulation is concerned.

EFFECT OF SPECTRAL ENVELOPE SMEARING ON SPEECH RECEPTION. II.*

ABSTRACT

This paper describes two experiments on the effect of reduced spectral contrast on the speech-reception threshold (SRT) for sentences in a background of interfering sound. Signal processing is performed by smoothing the envelope of the squared short-time fast Fourier transform by a convolution with a Gaussian-shaped filter, and overlapping additions to reconstruct a continuous signal. In the first experiment the effect of reduced spectral contrast on the SRT for male speech is investigated and compared with previously obtained results for female speech [chapter 2]. Spectral energy is smeared over bandwidths of 1/8, 1/4, 1/3, 1/2, 1, 2, and 4 oct. The results show that, despite the differences in spectral pattern between male and female voices, the SRT in noise increases similarly for both voices for smearing bandwidths over 1/3 oct. In terms of the ripple density of the spectral envelope the results indicate that the range of lower spectral modulations, up to a limit of about 1.5 periods/oct, is sufficient for the intelligibility of speech in interfering sounds. In the second experiment the extent of the threshold difference between a speech masker and a noise masker is investigated for spectral smearing bandwidths of 1/2, 1, and 2 oct. The release from masking found for the speech masker relative to the (steady-state) noise masker decreases with spectral envelope smearing.

**Paper accepted for publication In:* Journal of Acoustical Society of America (1992b)

INTRODUCTION

The perception of speech is a complex process which among other things depends on the ear's ability to resolve spectral features in the acoustic signal. This suggests that a reduction in the ear's ability to resolve spectral features, or a corresponding reduction in the signal's spectral features, will have a detrimental effect on speech intelligibility. With respect to spectral features, probably the most salient effect of limited frequency resolution is a blurring of the spectral envelope. Wider auditory filters will preserve less detail of the spectral envelope, thus making it more difficult to distinguish between the many spectrally based speech contrasts.

In a companion paper (chapter 2) we investigated the extent to which speech intelligibility and phoneme identification in a critical listening situation (background noise) depends on spectral contrasts. A reduction of spectral contrasts in the speech signal was used to simulate limited auditory spectral resolution. In this approach one assumes a reduction of spectral contrast in the speech signal to cause a loss of intelligibility by normal-hearing listeners in the same degree as though their auditory bandwidths were actually broadened. Reduction of spectral contrasts was accomplished by smearing the short-term spectral envelope of the speech signal on a log-frequency scale with a Gaussian-shaped filter. In this process spectral energy is smeared, but the signal's phase and harmonic structure are preserved. Results show that the speech-reception threshold (SRT) for sentences in noise, pronounced by a female speaker, increases as spectral energy is smeared over bandwidths exceeding 1/3 oct. Vowel identification suffers more from severe spectral envelope smearing than consonant identification. Vowels are primarily confused with the back vowels /ɔ,u/ and consonants are confused where place of articulation is concerned.

In the present study we continue our investigation of the extent to which the intelligibility of speech in a critical listening situation depends on spectral contrasts. This paper has two aims. One is to determine whether the intelligibility of male and female speech is differentially affected by spectral envelope smearing. It is unclear whether the differences between male and female speech reflected in the amount of spectral contrast in the speech signal, including position, amplitude and bandwidth of formant peaks and the depth of the valleys in between, are critical. In reducing spectral contrast, that is reducing the peak-to-valley differences in the short-term spectrum, formant slopes become shallower. This makes the detection of formant peaks, or any local spectral peak, relative to the surrounding components in a speech sound more difficult. As male speech (vowels) tends to have lower formant frequencies (in Dutch about 10%, Pols, 1977) than female speech (cf. Peterson and Barney, 1952; Fant, 1960; Childers and Wu, 1991), with the maxima in the spectral envelope generally being more closely spaced, male speech might be more susceptible to spectral envelope smearing than female speech. On the other hand, formant bandwidths of male speech tend to be narrower and formant slopes steeper (e.g. Fujimura and Lindqvist, 1971; Childers and Wu, 1991), which should improve discrimination (Horst, 1982). In running speech formants often do not reach the extreme values they attain when uttered in citation form, but it is unclear whether or how this affects the differences between male and female speech. Sometimes hearing-impaired listeners have a strong opinion on the differential intelligibility of male and female voices, often in favor of the female voice. Therefore, in the first experiment the SRT for sentences pronounced by a male speaker, as opposed to a female speaker in the previous paper, is measured in a background of steady-state noise as a function of the degree of spectral envelope smearing.

The second aim of this study is to determine whether the release from masking found for fluctuating maskers relative to steady-state maskers is affected by spectral envelope smearing. In a fluctuating interfering sound, for example, competing speech, normal-hearing listeners seem to benefit from the relatively silent intervals in the masker. Accordingly, the SRT for sentences in this condition is considerably lower than the SRT for sentences in steady-state noise. On the other hand, sensorineurally hearing-impaired listeners tend to show a severe reduction in this release from masking (e.g. Carhart and Tillman, 1970; Duquesnoy, 1983; Festen and Plomp, 1990). Several factors seem to play a role in this phenomenon, including reduced frequency resolution. To determine the contribution of spectral contrasts, and their resolution, to this phenomenon, the extent of the threshold difference between a noise and a speech masker is investigated in the second experiment for both male and female target speech for a number of smearing bandwidths.

I. METHOD

A. Signal processing

The signal processing consists of short-time fast Fourier transforms (FFT), low-pass filtering of the spectral envelope, and overlapping additions (OLA) to

reconstruct a continuous signal. Spectral contrasts are reduced by smoothing the short-term spectral envelope of the speech signal on a log-frequency scale with a Gaussian-shaped filter. The degree to which the spectral envelope is smeared is expressed in the equivalent rectangular bandwidth (ERB) of this filter. In this process spectral contrasts are gradually reduced, but the signal's phase and harmonic structure are preserved. The processing is described in detail in chapter 2.

B. Effect of spectral envelope smearing on the speech signal

The variation in spectral contrast in a signal may be quantified in terms of the ripple density (in periods per oct) present in the signal's short-term spectral envelope. Accordingly, smearing of the short-term spectral envelope may be quantified in terms of the ripples in periods/oct present in the envelope before and after smearing. The smearing process is accomplished by a convolution of the short-term spectral envelope on a log-frequency scale with a Gaussian-shaped function. A log-frequency scale with base 2 is chosen in order to be able to express the smearing bandwidth in oct. The smearing function is given by:

$$W(\log_2 f) = e^{-\pi\left(\frac{\log_2 f}{B}\right)^2} \qquad (1)$$

with $\log_2 f$: the logarithm of frequency with base 2, and B : the ERB of the smearing function in oct. The convolution is equivalent to a low-pass filtering of the spectral envelope with the Fourier transform of $W(\log_2 f)$, which is given by:

$$F\,[W(\log_2 f)] = B \, . \, e^{-\left(\frac{B^2}{4\pi}\right)d^2} \qquad (2)$$

with d : the ripple density in radians/oct. The cut-off ripple density of the envelope filter is defined as the density at which the spectral intensity ripples are reduced to half their original value. The cut-off ripple density can thus be computed:

$$e^{-\left(\frac{B^2}{4\pi}\right)d_{0.5}^2} = 0.5 \quad (3)$$

which gives:

$$d_{0.5} = \frac{2.95}{B}\ radians/oct \ \vee \ \frac{0.47}{B}\ periods/oct \quad (4)$$

Thus, the cut-off ripple density in the speech signal is approximately equal to the inverse of twice the bandwidth of the spectral-envelope smearing function.

C. Materials

The stimuli used in the two experiments - two sets of Dutch sentences - were digitized into 16 bits at a sampling frequency of 15,625 Hz. One set of sentences was pronounced by a female speaker (Plomp and Mimpen, 1979) and one set by a male speaker. For presentation the stimuli were converted to analog signals and low-pass filtered at 6250 Hz.

D. Subjects

Twenty-four normal-hearing university students between the ages of 19 and 30 served as paid listeners. Eight subjects participated in experiment 1, and sixteen in experiment 2. The stimuli in the two tests were presented monaurally to the right ear. All subjects had a pure-tone hearing level in the test ear lower than 15 dB over the frequency range from 125 Hz to 8 kHz, and were free of a history of ear pathology.

II. EXPERIMENT 1 : SPEECH RECEPTION IN NOISE

The aim of the present experiment is to determine the difference in intelligibility between male and female speech as a result of spectral envelope smearing. That is, whether differences between male and female speech reflected in the spectral contrasts in the speech signal, such as amplitude and bandwidth of formant peaks and the depth of the valleys in between, are critical. Therefore, the masked SRT for sentences pronounced by a male speaker, as opposed to a female speaker in the previous paper, is measured as a function of smearing bandwidth.

A. Stimuli

The speech material consisted of ten lists of 13 short (eight or nine syllables) sentences spoken by a male speaker. The masker was noise with the same spectrum as the long-term average spectrum of the sentences.

Eight conditions were investigated : the signal analyzed and resynthesized by the FFT-OLA system but unsmeared (US), and the signal in seven conditions of spectral smearing with ERB 1/8, 1/4, 1/3, 1/2, 1, 2, and 4 oct, respectively. The corresponding range of spectral ripple densities in the speech signals varies from 4 periods/oct to 1/8 period/oct. The first mentioned ripple density agrees with the estimated upper limit of spectral ripple densities the ear is able to detect when the effect of lateral suppression is included (Houtgast, 1974; Plomp, 1983). Without this effect the upper limit is estimated to be about 2 periods/oct (smearing bandwidth of 1/4 oct).

The spectral-smearing process changes the short-term spectral slope towards a straight line with -3 dB/oct. Because the average speech spectrum is different, the process changes to some degree the overall slope of this spectrum. In order to preclude extra effects of masking or unmasking, both speech and noise were subjected to the same degree of spectral smearing in each condition. Speech and noise were processed separately in order to obtain a degree of spectral smearing independent of the signal-to-noise ratio in the mixed signal.

B. Test procedure

Eight subjects listened to the lists of sentences monaurally (right ear) through headphone (Beyer DT 48) in a soundproof room. The lists were presented in fixed order. The sequence of the smearing conditions applied was varied over listeners according to a 8x8 digram-balanced Latin square to minimize effects of the order of presentation of the conditions, and effects of sentence lists in the average results. Prior to the administration of the test material the listeners were presented with one list of sentences spectrally smeared over 1 oct and another list smeared over 4 oct to familiarize them with the quality of the speech and with the experimental task.

The masked SRT is defined as the speech-to-noise ratio in decibels at which 50% of the sentences can be reproduced without a single error. The level of the masking noise was fixed at 65 dBA and the level of the sentences was changed according to an up-down adaptive procedure. In the adaptive procedure, the first sentence of each list is initially presented at a level below the reception threshold. This sentence is then presented repeatedly, each time at a 4-dB higher level, until the listener is able to reproduce it correctly. The remaining 12 sentences in a list are presented only once in a simple up-down procedure with a step size of 2 dB. An errorless reproduction of the entire sentence is required for a correct response. The average presentation level over sentences 4-13 is taken as the SRT.

C. Results and discussion

Figure 1 shows the mean masked SRT for sentences pronounced by a male speaker as a function of the degree of spectral envelope smearing. Critical differences between smearing conditions were evaluated with an analysis of variance for a repeated-measures Latin-square design (Hays, 1988). A significant effect of *spectral envelope smearing* was found ($F=61.57$, $p<0.0001$). Post-hoc analysis (Tukey HSD, Neter *et al.*, 1990) revealed a significant difference ($p<0.001$) between the SRT in the unsmeared condition and the SRT for spectral smearing over 1/2 oct and upward. The SRT for unsmeared speech, and for spectral smearing over 1/8, 1/4, and 1/3 oct are not statistically different, and the threshold is on average -4.0 dB.

The primary aim of the present experiment was to determine the difference in effectiveness of spectral envelope smearing on the intelligibility of male and female speech. Therefore, the present results are compared with the results obtained earlier for the same type of sentences in similar smearing conditions, but pronounced by a female speaker (chapter 2). A repeated-measures analysis of variance was performed on the pooled data with one between-subjects factor (*female versus male speaker*) and one within-subjects factor (*smearing condition*). This analysis showed: (1) a highly significant effect of *spectral envelope smearing*, with SRT increasing as a function of increasing smearing bandwidth ($F=108.19$, $p<0.0001$); (2) no effect of

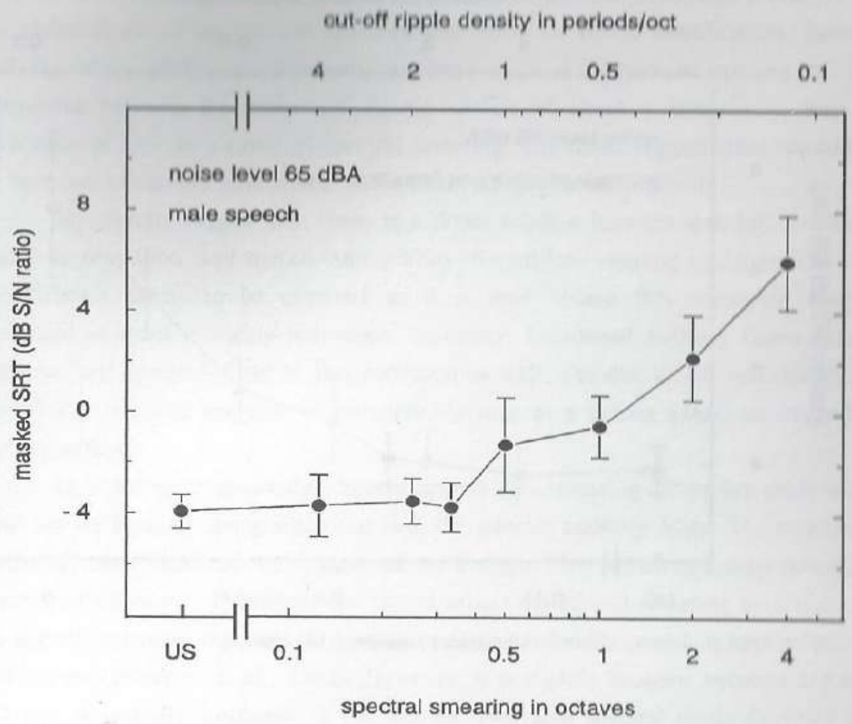


FIG.1. Mean speech-reception threshold for sentences in noise as pronounced by a male speaker, expressed in signal-to-noise ratio, as a function of smearing bandwidth. Corresponding cut-off ripple densities are shown along the top horizontal axis. Vertical bars represent the standard deviation of the SRT in each condition.

different speakers, and (3) no interaction between *speaker* and *smearing conditions*. Again, post-hoc analysis (Tukey HSD) indicated that the SRT for smearing conditions of 1/2 oct and upward differs significantly ($p<0.001$) from the SRT found in the unsmeared condition. The SRT resulting from the pooled data is given in Fig. 2.

For both male and female speech the SRT for sentences in noise increases significantly for smearing bandwidths over 1/3 oct. In terms of ripple densities in the spectral envelope this indicates that the lower (spectral) ripple densities up to a limit of about 1.5 periods/oct are sufficient for the intelligibility of speech in a background of interfering noise. A comparable critical (spectral) ripple density was found

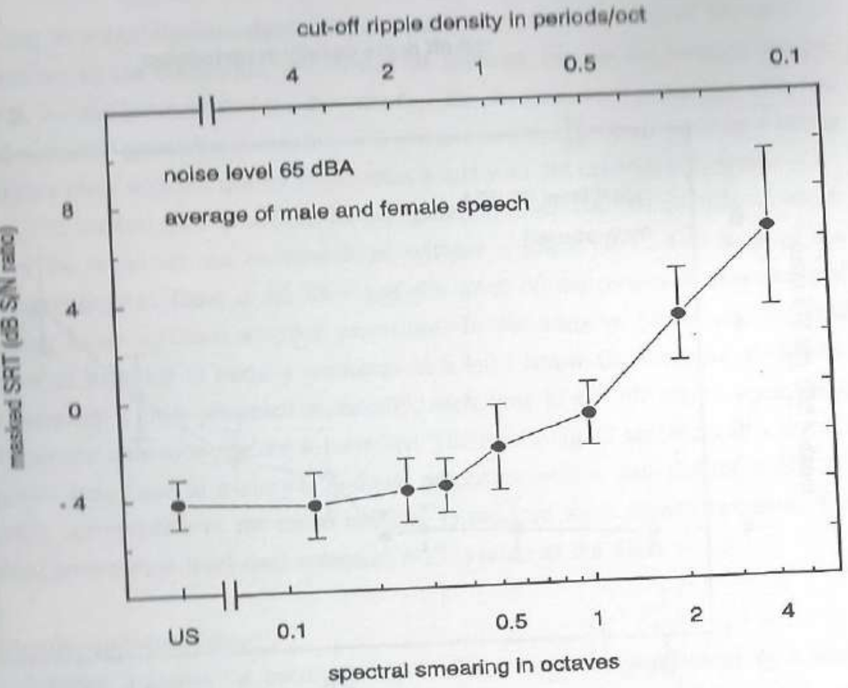


FIG. 2. Speech-reception threshold for sentences in noise averaged over a male and a female speaker as a function of smearing bandwidth. Corresponding cut-off ripple densities are shown along the top horizontal axis. Vertical bars represent the standard deviation of the SRT in each condition.

in masked consonant identification (chapter 2). These findings are in agreement withresults by van Veen and Houtgast (1985), who found that the dissimilarities between synthetic vowels are determined mainly by the global shape of the spectral envelopes, i.e., by spectral ripples up to about 1.5-2 periods/oct. More detailed variation in the spectral envelope appeared to be of minor influence on vowel dissimilarities. In turn, this critical ripple density agrees with the estimated upper limit of spectral ripple densities the ear is able to detect when the effect of lateral suppression is not included (Houtgast, 1974; Plomp, 1983). The results suggest that the detailed differences in spectral pattern between male and female speech are not

critical for the intelligibility of speech, as intelligibility seems to depend primarily on the global shape of the spectral envelope (see also, for vowel identification, Beddor and Hawkins, 1990). Furthermore, as there is a difference in spacing of the harmonics between the male and female speech of about a factor two, but no difference in SRT as a result of spectral smearing, this result suggests that resolution of harmonics does not have a noticeable effect on speech intelligibility.

The results suggest that there is a direct relation between spectral contrasts, and their resolution, and speech intelligibility in a critical listening situation. The last specification seems to be essential as it is well known that connected speech presented in quiet is highly redundant. Evidently, broadened auditory filters would preserve less spectral detail in this condition as well, but due to the redundancy of speech the resulting decrease in intelligibility may to a certain extent be overruled (ceiling effects).

In listening to spectrally smeared speech two smearing filters are dealt with. One, the external smearing filter and two, the internal auditory filter. The combined action of these filters can be represented by a single filter which is a convolution of the individual filters. Therefore, one would expect the critical smearing bandwidth to be slightly narrower than the ear's auditory filter bandwidth, which is approximately 1/4-1/3 oct (Patterson *et al.*, 1982). However, it is slightly broader: between 1/3 and 1/2 oct. A possible conclusion is that not all resolvable spectral ripple densities are critical for the intelligibility of speech, as indicated by the findings that intelligibility depends primarily on the global shape of the spectral envelope. Together, this may explain the findings in several studies that vowel identification in quiet and the minimal detectable depth of a spectral notch sufficient for vowel identification seem to be relatively insensitive to variations in selectivity (e.g. Klein et al., 1970; Turner and Van Tasell, 1984).

III. EXPERIMENT 2 : SPEECH RECEPTION IN INTERFERING SPEECH

The aim of the second experiment is to determine whether the release from masking found for fluctuating maskers relative to steady-state maskers is reduced by spectral envelope smearing. Therefore, the difference in SRT obtained with a noise masker and a speech masker is investigated for a number of smearing bandwidths for both male and female target speech.

A. Stimuli

The speech material consisted of ten lists of 13 sentences pronounced by a male speaker, as used in experiment 1, and ten comparable lists of 13 short (eight or nine syllables) sentences pronounced by a female speaker. As masker, both pseudo-running speech and noise, with a spectrum equal to the long-term average spectrum of the sentences were used. Pseudo-running speech was obtained by generating various sentences in succession without pauses. Target sentences of the female speaker were masked by pseudo-running speech of the male speaker or noise with that spectrum and vice versa.

Four conditions of spectral smearing were investigated: the signal analyzed and resynthesized by the FFT-OLA system but unsmeared (US), and the signal smeared over 1/2, 1, and 2 oct. Apart from the overall difference in long-term average spectrum between the male and the female speaker, small variations in the *average* spectrum that are introduced by the spectral smearing should be present both in the target speech and in the masker (pseudo-running speech and noise) to preclude unwanted masking. Therefore, both speech and masker were subjected to the same degree of spectral smearing in each condition. Speech and noise were processed separately for the same reasons as in experiment 1.

B. Test procedure

Apparatus and stimulus-presentation paradigm were similar to those used in experiment 1. The sentences were presented monaurally (right ear) through headphone (Sony MDR-CD999) in a soundproof room. Each of 16 subjects listened to both the male and the female target sentences in eight conditions (4 smearing conditions x 2 masker conditions). Eight listeners started with the female target speech in eight different conditions and eight listeners with the male target speech. The lists were presented in fixed order. The sequence of presentation of the eight conditions was varied over listeners and lists according to a 8x8 digram-balanced Latin square to minimize order and list effects in the average results. Prior to the administration of the tests with the male and the female speaker the listeners were presented with one list of sentences smeared over 1/2 oct and another list smeared over 2 oct to familiarize them with the quality of the speech and the experimental task. The level of the masker was fixed at 65 dBA and the level of the target sentences was changed according to an up-down adaptive procedure. Once again, the masked SRT is defined as the speech-to-masker ratio in decibels at which 50% of

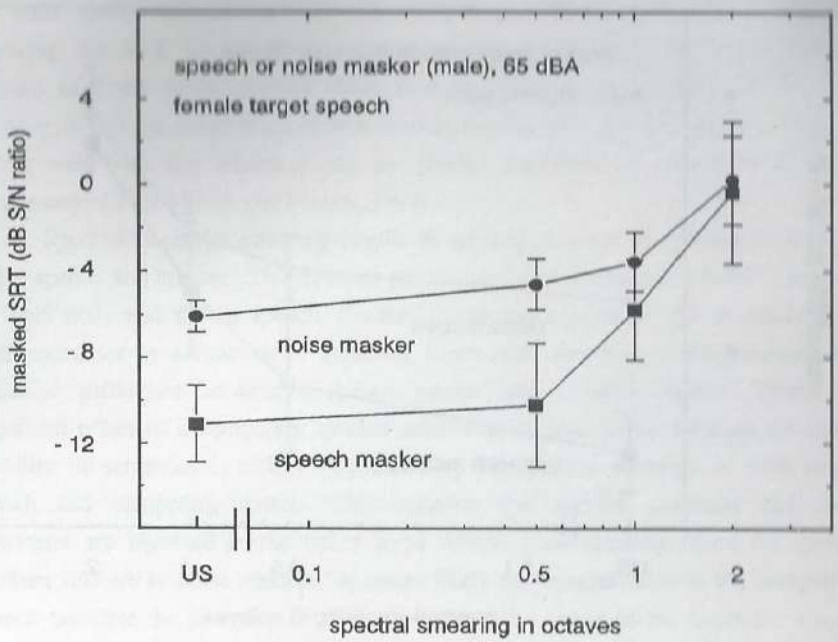


FIG. 3. Mean speech-reception threshold for female target speech in male competing speech and noise as a function of smearing bandwidth. Vertical bars represent the standard deviation of the SRT in each condition.

the sentences can be reproduced without a single error.

C. Results and discussion

Figures 3 and 4 show the mean SRT for sentences in noise and in competing speech as a function of smearing bandwidth. Figure 3 shows the results for female target speech masked by male speech or noise with a corresponding spectrum, and Fig. 4 shows the results for male target speech masked by female speech or noise. A repeated-measure analysis of variance on the data showed four highly significant effects : (1) an effect of *spectral envelope smearing*, with SRT increasing as a function of increasing smearing bandwidth (F=239.52, $p<0.0001$); (2) an effect of *masker type*, with the speech masker giving a lower SRT than the steady-state noise masker (F=173.63, $p<0.0001$); (3) an interaction of *masker type* and *spectral*

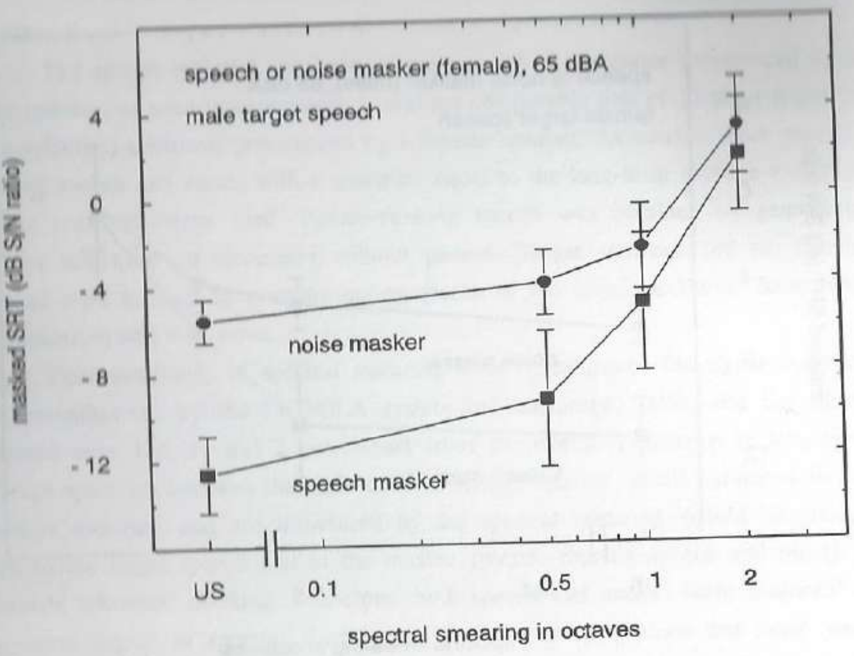


FIG. 4. Mean speech-reception threshold for male target speech in female competing speech and noise as a function of smearing bandwidth. Vertical bars represent the standard deviation of the SRT in each condition.

envelope smearing, with the difference in SRT between speech and noise masker decreasing with increasing smearing bandwidth (F=18.73, p<0.0001); (4) an effect of *target speech*, with the SRT for male target speech being slightly higher than the SRT for female target speech (F=10.76, p<0.001).

The effect of target speech is probably an effect of the overall difference in spectrum between the two voices. The male voice seems to be more adequately masked by a sound with the spectrum of the female voice than the other way around. Spectral components above 3 kHz are up to about 10 dB stronger in the female voice than in the male voice. Additionally, voice-related effects, like clearness of articulation and speech rate, may play a role.

For the female speech in the unsmeared condition the mean SRT in noise is -6 dB, and -11 dB in competing speech. In the same conditions the mean SRT for

the male speech is -5.5 dB and -12.6 dB, respectively. Thus, without spectral smearing the SRT for sentences masked by competing speech is 5 dB (female stimuli) to 7 dB (male stimuli) lower than the SRT in noise. This release from masking of 5-7 dB found for a speech masker relative to a steady-state noise masker agrees well with the release found for similar conditions in other studies (e.g. Duquesnoy, 1983; Festen and Plomp, 1990).

Spectral envelope smearing results in spectral contrast being reduced in both target speech and masker. This reduces the intelligibility of the target speech: that is, for both male and female speech, the SRT in competing speech and in steady-state noise increases as a function of smearing bandwidth. However, it also reduces the threshold difference between a speech masker and a noise masker. Thus, the beneficial effect of a competing speaker relative to an interfering noise on the intelligibility of sentences is reduced by smearing the spectral envelope of both target speech and competing speech. This suggests that spectral contrasts and their resolution are involved in the rather large release from masking found for speech maskers relative to noise maskers. It seems likely that spectral dips in the competing speech facilitate the detection of the target speech by virtue of the favorable signal-to-noise masker ratio in these regions. This suggests that the extent of the difference in SRT between competing speech and noise should decrease as a function of spectral envelope smearing. Likewise, Festen and Plomp (1990) found that the threshold difference between a speech masker and a steady-state noise masker decreases, when, instead of competing speech, noise modulated with the broad-band intensity fluctuations of speech was used as masker. Additionally, although the gross temporal structure of the signal is preserved by spectral envelope smearing, variations in a limited frequency region in the temporal envelope may be introduced due to the variation of the spectral modulations over time. This may also contribute to the decrease in threshold difference.

IV. SUMMARY AND CONCLUSIONS

In the present study we investigated the extent to which reduced spectral contrast affects the intelligibility of speech in a critical listening situation. Spectral contrasts were reduced by smearing the short-term spectral envelope of the speech signal on a log-frequency scale with a Gaussian-shaped filter. The degree to which the spectral envelope is smeared is expressed in the bandwidth of this filter.

In the first experiment the effect of reduced spectral contrast on the SRT for male versus female speech in noise was investigated. Results show that the intelligibility of male and female speech is not differentially affected by spectral envelope smearing, in spite of detailed differences in spectral pattern. For both male and female speech, the SRT increases for smearing bandwidths exceeding 1/3 oct. This result is a clear indication of the detrimental effect of limited resolution of spectral contrasts on the intelligibility of speech in a critical listening situation. In terms of ripple densities of the short-term spectral envelope, the lower spectral ripples, up to a limit of about 1.5 periods/oct, seem to be sufficiant for the intelligibility of speech in noise. This indicates that the intelligibility depends primarily on the global shape of the spectral envelope.

In the second experiment the effect of reduced spectral contrast on the threshold difference between speech and noise masker was investigated. Results show that this threshold difference decreases with increasing smearing bandwidth. This indicates that spectral contrasts, and their resolution, are involved in the rather large release from masking found for the intelligibility of speech masked by competing speech relative to speech masked by steady-state noise.

LIMITED RESOLUTION OF SPECTRAL CONTRAST AND HEARING LOSS FOR SPEECH IN NOISE.*

ABSTRACT

This paper examines the relation between the spectral contrasts needed in speech discrimination and hearing loss for speech in noise. Fifteen hearing-impaired listeners with relatively flat, mild-to-moderate sensorineural losses and eight normal-hearing listeners participated in the study. As a criterion of the spectral contrasts essential to speech discrimination the reduction in contrasts in speech beyond which the speech-reception threshold (SRT) for sentences in noise increases was used. Reduction of spectral contrast was accomplished by smearing the envelope of the squared short-time fast Fourier transform by a convolution with a Gaussian-shaped filter, and overlapping additions to reconstruct a continuous signal. Auditory filter bandwidth was determined by estimating auditory filter shapes at center frequencies, 0.8, 1.6, and 3.2 kHz, using a notched-noise masking paradigm. Correlation coefficients were computed between the various findings. The results show that limited resolution of spectral contrast is only loosely associated with hearing loss for speech in noise. Moreover, the correlations between the SRT for unsmeared speech and the auditory filter bandwidth at various frequencies were weak.

**Paper submitted for publication in:* Journal of the Acoustical Society of America (1992c)

INTRODUCTION

Limited frequency resolution has often been shown to correlate with hearing loss for speech in noise (e.g., Leshowitz, 1977; Bonding, 1979; Horst, 1987; Festen and Plomp, 1983; Dreschler and Plomp, 1985; Stelmachowitz *et al.*, 1985), but a direct relation has proven difficult to establish. Generally, limited frequency resolution is thought to affect speech intelligibility in noise both because broadened filters will pass more background noise, making it more difficult to detect the speech signal, and because broadened auditory filters will preserve less spectral detail, making it more difficult to distinguish between the many spectrally based speech contrasts.

To quantify the importance of spectral contrasts to speech intelligibility we investigated, in two previous papers with normal-hearing listeners, the extent to which speech intelligibility in a critical listening situation is affected by a reduction in spectral contrasts (chapters 2 and 3). Reduction of spectral contrast was accomplished by smearing the short-term spectral envelope of the speech signal on a log-frequency scale with a Gaussian-shaped filter. In this process spectral energy is smeared, but the signal's phase and harmonic structure are preserved. The signal processing may be interpreted as the effect of limited resolution of the power-spectrum envelope isolated from auditory processing. The results show that sentence intelligibility decreases progressively for smearing over frequency bands wider than 1/3 oct. The smearing bandpass filter can also be interpreted as a low-pass filter of the ripple densities in the spectral envelope. In these terms, the results from the previous two papers indicate that speech intelligibility in a critical listening situation depends primarily on the global shape of the spectral envelope. That is, the range of lower ripple densities in the spectral envelope up to about a limit of 1½ periods/oct appears to be sufficient. Although these results establish a direct relation between spectral contrasts and speech intelligibility in noise, it is still unclear whether limited resolution of spectral contrasts may explain (part of) hearing loss for speech in noise.

The primary aim of the present study, therefore, is to examine for sensorineurally hearing-impaired listeners the relation between the spectral contrasts needed in speech discrimination and hearing loss for speech in noise.

In the first experiment, the upper limit of spectral contrasts needed in speech discrimination is determined by measuring the smearing bandwidth beyond which the speech-reception threshold (SRT) for sentences in noise increases, i.e. the critical smearing bandwidth (CSB). The SRT for unsmeared speech as a measure of a listener's hearing loss for speech in noise is taken as a baseline. In the second experiment, auditory filter bandwidth is measured by estimating auditory filter characteristics with a notched-noise paradigm in simultaneous masking (Patterson, 1976). Finally, the relations between the various findings are discussed.

I. METHOD

A. Subjects

Fifteen sensorineurally hearing-impaired (HI) listeners between the ages of 22 and 65 years, and eight normal-hearing (NH) university students between the ages of 18 and 23 years served as paid listeners. HI listeners with relatively flat audiograms were chosen, since the smearing process is based on spectral contrasts being reduced to the same degree over the whole frequency range, and limited frequency resolution is typically, though not always, confined to the region of absolute threshold loss (cf. Tyler, 1986). The HI listeners showed mild to moderate sensorineural hearing losses (25-65 dB) over the frequency range 250 - 6000 Hz and were tested at the ear with the smallest average loss. Table I shows for the HI listeners the absolute threshold for pure tones at the test ear, and other relevant data. HI listeners were selected to exclude the involvement of conduction loss: air-bone gaps did not exceed 10 dB. Identification of monosyllables in quiet reached at least 80%. The latter restriction was imposed, since the speech-reception threshold for sentences in noise is scored in terms of the signal-to-noise ratio required for 50% correct identification of complete sentences. All NH listeners showed a pure-tone hearing level in the test ear (left) lower than 15 dB over the frequency range 250 - 6000 Hz, and were free of a history of ear pathology.

B. Signal processing to reduce spectral contrast

The signal processing to reduce spectral contrasts consists of fast Fourier transforms (FFT) on 16.4-ms frames, spectral processing, inverse Fourier transforms, and overlapping additions (OLA) to reconstruct a continuous signal. Spectral contrasts are reduced by smoothing the short-term spectral envelope of the speech signal on a log-frequency scale with a Gaussian-shaped filter. The degree to which

TABLE I. Audiometric data for the 15 sensorineurally hearing-impaired listeners. For each listener, the table shows age, sex, test ear, and the absolute threshold in dB HL at nine frequencies in kHz.

Subj	Age	Sex	Ear	Threshold, dB HL								
				0.25	0.5	0.8	1.0	1.6	2.0	3.2	4.0	6.0
1	53	F	L	20	25	25	30	40	40	40	40	35
2	22	F	R	60	65	65	65	60	55	50	50	50
3	65	M	L	40	55	55	60	55	50	55	60	55
4	64	F	L	30	40	40	35	30	40	30	40	40
5	59	F	L	30	30	35	40	35	35	45	50	50
6	48	F	R	35	35	40	45	40	35	35	35	40
7	34	F	L	25	30	35	40	50	50	50	50	50
8	36	M	L	35	30	40	50	50	50	55	55	75
9	23	F	R	45	50	55	60	60	60	50	60	60
10	46	F	L	40	45	50	50	55	55	50	60	60
11	64	M	L	50	50	50	50	55	55	55	55	50
12	50	F	L	40	40	30	30	35	35	35	35	40
13	43	F	L	45	55	55	55	60	60	55	50	55
14	46	M	L	50	50	50	55	55	55	50	50	35
15	38	F	R	65	65	65	60	65	65	65	65	70

the spectral envelope is smeared is expressed in the equivalent rectangular bandwidth (ERB) of this filter. In this process, spectral contrasts are gradually reduced, while the signal's phase and harmonic structure are preserved. The processing is described in detail in chapter 2.

II. EXPERIMENT I: SPECTRAL CONTRASTS NEEDED IN SPEECH RECEPTION

A. Stimuli

The speech material consisted of ten lists of 13 short (eight or nine syllables) Dutch sentences (Plomp and Mimpen, 1979). The stimuli were digitized into 16 bits at a sampling frequency of 15,625 Hz. All stimuli were pronounced by a female speaker. For presentation, the stimuli were converted to analog signals, and low-pass filtered at 6250 Hz. The masker was noise with the same spectrum as the long-term average spectrum of the sentences.

Eight conditions were investigated: the signal analyzed and resynthesized by the FFT-OLA system but unsmeared (US), and the signal in seven conditions of spectral smearing with ERB 1/3, 1/2, 1, 1.5, 2, 3, and 4 oct, respectively. In order to preclude overall spectral differences between signal and masker, both speech and noise were subjected to the same degree of spectral smearing in each condition. Speech and noise were processed separately in order to obtain a degree of spectral smearing independent of the signal-to-noise ratio in the mixed signal.

B. Procedure

All subjects (fifteen HI and eight NH listeners) listened to the sentences monaurally through headphone (Sony MDR-CD999) in a soundproof room. In this study the prime interest is not the average speech-reception threshold as a function of spectral-envelope smearing, but the relation between the SRT and the bandwidth of the auditory filter. For this reason sentence lists and smearing conditions were presented to the subjects in a fixed order: US, 1/3, 1/2, 1, 1.5, 2, 3, and 4 oct, respectively. In this way, fatigue and learning effects have a minimal effect on the correlations at the expense of an effect on the average results. Prior to the administration of the test, the listeners were presented with a list of sentences spectrally smeared over 1 oct and another list smeared over 4 oct to familiarize them with the quality of the speech and with the experimental task.

The masked SRT is defined as the speech-to-noise ratio in decibels at which 50% of the sentences can be reproduced without a single error. The SRT is measured by changing the level of the sentences relative to the level of the masker according to an up-down adaptive procedure. In this adaptive procedure, the first sentence of each list is presented at a level well below the level of the masker. This sentence is then presented repeatedly, each time at a 4-dB higher level, until the listener is able to reproduce it correctly. The remaining twelve sentences in a list are presented only once in a simple up-down procedure with a step size of 2 dB. An errorless reproduction of the entire sentence is required for a correct response. The average presentation level adjusted after sentences 4-13 is taken as the SRT for that particular condition.

TABLE II. Masker levels for the 15 hearing-impaired listeners. For each listener, the masker level in dBA used in the SRT experiment and the masker spectrum levels in dB/Hz used in the notched-noise experiment are shown.

	Masker level	Spectrum level notch noise dB/Hz		
Subj	SRT dBA	800 Hz	1600 Hz	3200 Hz
1	85	52	46	40
2	95	62	56	50
3	90	57	51	45
4	85	52	46	40
5	85	52	46	40
6	85	52	46	40
7	95	62	56	50
8	95	62	56	50
9	95	62	56	50
10	95	62	56	50
11	95	62	56	50
12	85	52	46	40
13	95	62	56	50
14	95	62	56	50
15	95	62	56	50

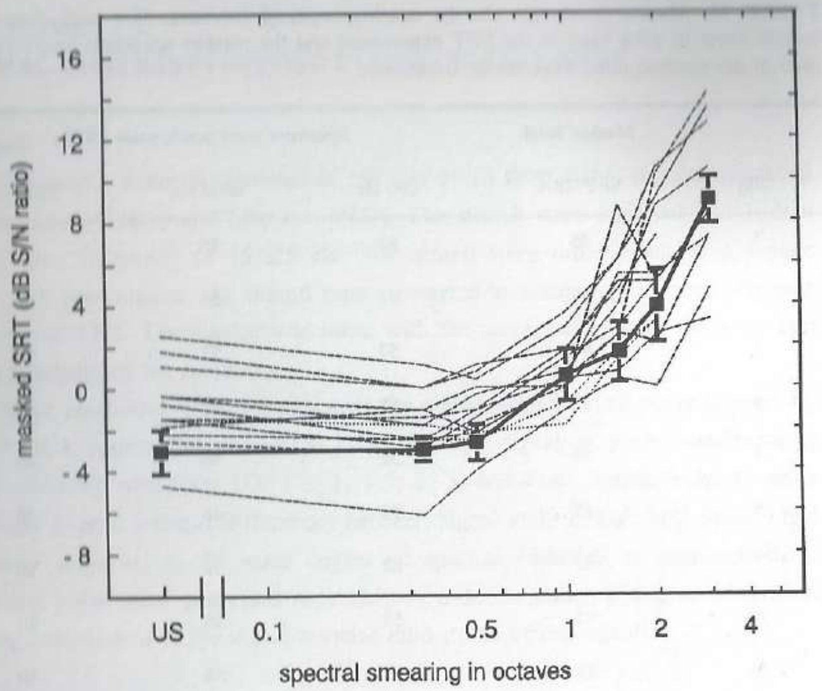


FIG. 1. Speech-reception threshold for sentences in noise, expressed in signal-to-noise ratio, as a function of the bandwidth of spectral envelope smearing. In each condition, the mean SRT and corresponding standard deviation are given for the normal-hearing listeners; individual data are given for the hearing-impaired listeners.

For each HI listener, the level of the noise masker was fixed at a level which was intermediate between the listener's absolute threshold for pure tones and the listener's level of uncomfortable loudness over the frequency range 250-4000 Hz. Table II shows for the HI listeners the noise levels used: 85, 90, and 95 dBA. With 1/3 oct bands of noise it was ensured that the masker was presented at least 10 dB above the listener's absolute threshold over the frequency range 250 - 4000 Hz. For the NH listeners, the masker level was fixed at 85 dBA.

C. Results and discussion

Figure 1 shows the SRT for sentences in noise as a function of the bandwidth of spectral envelope smearing. Mean thresholds are given for the NH listeners,

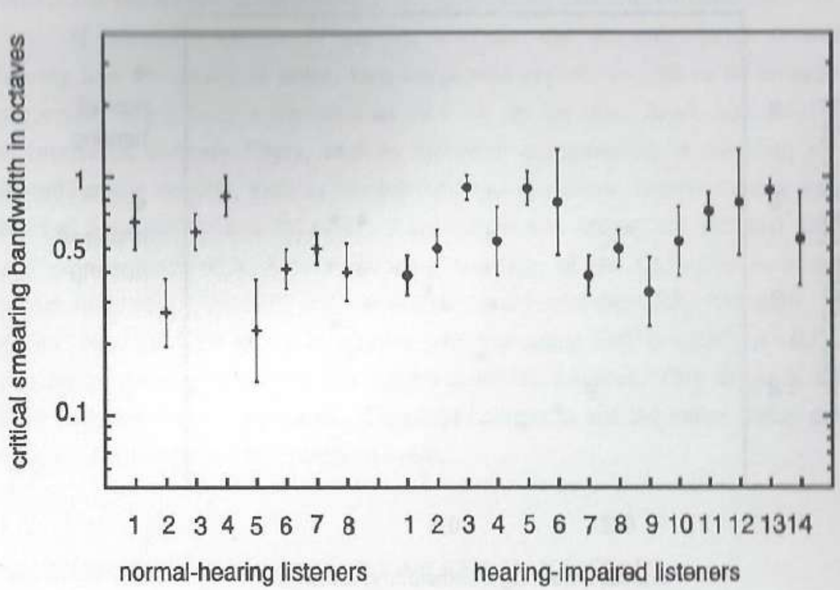


FIG. 2. Fitted critical smearing bandwidth in oct for eight normal-hearing and fourteen hearing-impaired listeners. Vertical bars represent the standard deviation of the fitted bandwidth.

individual thresholds for the HI listeners. HI subject 15 experienced such severe problems in understanding speech in noise that for none of the conditions a reliable SRT could be determined. Furthermore, it appeared impossible to determine a reliable SRT in the 4-oct condition for all HI listeners, and for some HI listeners in the 3-oct condition. Therefore, data from the 3- and 4-oct condition of spectral smearing were excluded from statistical analysis.

The NH listeners show speech-reception thresholds that are somewhat higher than the thresholds obtained for the same speaker in similar conditions in a previous study (chapter 2). This may have been a result of the fixed order of presentation of the smearing conditions and sentence lists. Nevertheless, as this effect is constant across listeners it will not influence the comparison between listeners or the correlations to be computed. The HI listeners show curves of the same shape as those of the NH listeners, but many are elevated, which points to an overall loss in intelligibility irrespective of spectral smearing. As can be seen from Figure 1 there is wide

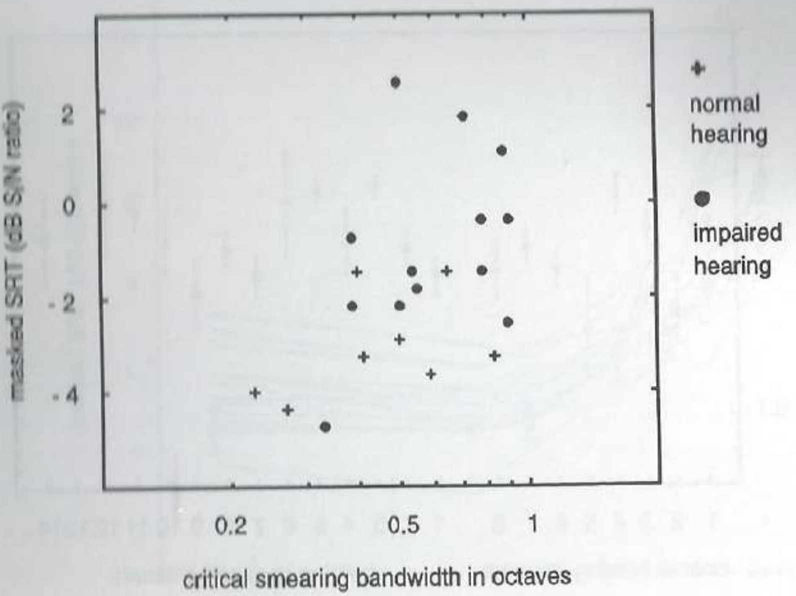


FIG. 3. Speech-reception threshold for unsmeared speech in noise for eight normal-hearing and fourteen hearing-impaired listeners plotted against their critical smearing bandwidth.

variation in hearing loss for speech in noise across HI listeners: thresholds for unsmeared speech vary from about -3 dB, the average SRT of the NH listeners in this condition, to +2 dB. Surprisingly, one HI listener actually showed better-than-normal speech intelligibility in all conditions (i.e., lower SRTs).

The critical smearing bandwidth was determined for each NH and HI listener using a nonlinear regression technique (Dixon, 1985, module 3r). The SRT for unsmeared speech, and for smearing conditions 1/3, 1/2, 1, 1.5, and 2 oct, were fitted on a log-bandwidth scale with two straight line segments, one of which was kept horizontal, in an iterative procedure which estimates the fit with the smallest residual sum of squares. The CSB is defined as the bandwidth in oct at the intersection of the two fitted line segments. The residual sum of squares is used in computing the standard deviation of the fitted line segments and inflexion bandwidth.

Figure 2 shows for each NH and HI listener the fitted CMB and the standard deviation of the fit. Most remarkably, there is no systematic difference in CSB between NH and HI listeners. The CSB varies from about 1/4 - 1 oct among listeners, but the spread in bandwidth is comparable for NH and HI listeners.

If limited resolution of spectral contrast was the only factor determining hearing loss for speech in noise, then one would expect the CSB to be broadened in proportion to a listener's elevation in SRT. If, on the other hand, additional effects of broadened auditory filters, such as increased susceptability to masking, or other suprathreshold deficits, such as limited temporal resolution, contribute to a listener's hearing loss for speech in noise, then one would expect an elevated SRT for unsmeared speech without a proportional widening of the CSB. Figure 3 shows a scatter diagram of the SRT for unsmeared speech and the CSB. Although, for the pooled data, the CSB seems to broaden with increasing SRT (r=0.45, $p<0.02$), the relation is non-significant for the subgroup of HI listeners. This suggests that for these listeners limited resolution of spectral contrast is not the major factor determining their hearing loss for speech in noise.

III. EXPERIMENT 2: AUDITORY FILTER BANDWIDTH

To determine a listener's ability to resolve frequency, auditory filter shapes were estimated at center frequencies 0.8, 1.6, and 3.2 kHz. Filter characteristics were estimated by measuring the threshold for a sinusoidal signal in a notched-noise masker (Patterson, 1976). The notch was positioned both symmetrically and asymmetrically about the signal frequency on a linear frequency scale. The signal and masker were presented simultaneously, and signal threshold was measured as a function of the width of the notch in the noise.

Several studies have suggested that frequency resolution may depend to some degree on stimulus level (e.g., Pick, 1980; Lutfi and Patterson, 1984). To avoid a confounding effect of stimulus level when comparing speech-reception abilities with frequency-resolution abilities, the presentation level of the notched-noise stimuli was adjusted for each listener to the presentation level of the masker in the speech-reception experiment.

A. Stimuli

The signal frequency, f_s, was a sinusoid of 0.8, 1.6, or 3.2 kHz, presented for 400 ms including cosine rise and fall times of 50 ms. The notch masker was produced by positioning two bands of noise with very steep skirts and a flat top symmetrically or asymmetrically about the signal on a linear frequency scale. The complete masker (the sum of the two noise bands) came on 50 ms before the signal and terminated 50 ms thereafter.

Noise bands were prepared by defining the spectrum in a 8192-point array and performing an inverse Fourier transform to obtain a 524-ms noise fragment with 64 μs intersample time. Within each noise band the amplitude of all Fourier components (spacing 1.9 Hz) was fixed and the phase angles were random. The width of each noise band was $0.4f_s$. In each trial the masker was randomly chosen from a set of 50 maskers with different random phases.

The deviation of the nearer edges of the upper or lower noise bands from the signal frequency f_s is denoted as Δf. In symmetrical conditions $\Delta f/f_s$ was: 0.0, 0.2, 0.4, and 0.6. In the asymmetrical conditions the near edge of the upper or lower masker bands was $0.2f_s$ farther away from the signal frequency. In the upper and lower asymmetrical conditions the relative deviations from the signal frequency were: 0.2/0.4, 0.4/0.6, 0.6/0.8, and 0.4/0.2, 0.6/0.4, 0.8/0.6, respectively. The width of the lower frequency band was reduced where necessary to avoid "wrapping around" zero frequency.

B. Procedure

The subjects listened to the stimuli monaurally through headphone (Sony MDR-CD999) in a soundproof room. Masked thresholds were measured using a two-alternative forced-choise adaptive procedure that estimates the 70.7% point on the psychometric function (Levitt, 1971). Measurements at the three center frequencies were performed in fixed order: 1.6, 0.8, and 3.2 kHz. Notch conditions for each center frequency were completed in random order.

For each combination of center frequency and notch width, the threshold was determined in the following way. In the first trial, the signal was presented 15 dB above the expected threshold level. The signal level was decreased after two consecutive correct responses and increased after each incorrect response, with an initial step size of 4 dB. After three reversals in level (the transition from decreasing to increasing level, and vice versa) the step size was reduced to 2 dB. Testing continued until 12 reversals occurred, and the average level of the last 8 reversals was taken as the threshold. The standard deviation of the reversals was typically less than 5 dB.

Each trial consisted of two bursts of the masker, and in one of them, chosen

randomly, the signal was presented. The listener was instructed to press a button corresponding with the interval containing the signal. When the correct interval was chosen, feedback was provided by a red light.

For each listener and each center frequency, a spectrum level for the noise masker was chosen that approximated the level of the speech-shaped noise masker in the corresponding frequency region in the SRT experiment. The spectrum levels chosen for the HI listeners are shown in Table II. Spectrum levels for the NH listeners were 52, 46, and 40 dB/Hz at the respective center frequencies of 0.8, 1.6, and 3.2 kHz.

C. Results and discussion

Auditory-filter characteristics were derived from the notched-noise data assuming that each side of the auditory filter has the form of the roex(p,r) filter described by Patterson *et al.* (1982). In this model, p is a parameter determining the slope of the filter skirt (values of the lower and upper side are denoted by p_l and p_u, respectively), and r is a parameter that delimits the dynamic range of the filter. It is assumed that r has the same value for both sides of the filter. The fitting procedure is based on a program published by Glasberg and Moore (1990), and includes two of their modifications. That is, changes in filter bandwidth with center frequency when allowing for the effects of off-frequency listening are taken into account, and the amount by which the center frequency of the filter can shift in order to maximise the signal-to-masker ratio is limited. From this fitting procedure the equivalent rectangular bandwidth (ERB) is taken as a measure of the selectivity of the filter.

Parameters of the fitted auditory filters are given in Table III for the NH listeners, and in Tables IV and V for the HI listeners. Masked thresholds for HI listener 15 were more or less constant as a function of notch width. Therefore, no reliable filter shapes could be determined for this listener. Consequently, this subject was, just as in the speech-reception experiment, excluded from further analysis.

An analysis of variance on the fitted filter parameters indicates that the group of HI listeners generally shows poorer frequency resolution on all auditory-filter characteristics. That is, (1) the *ERB* tends to be broader (F=51.372, $p<0.0001$); (2) consequently, as p is inversely related to the ERB, the slopes, p_l and p_u, tend to be shallower (F=15.562, $p<0.0001$; F=25.633, $p<0.0001$, respectively); and (3), the dynamic range r tends to be reduced (F=146.509, $p<0.0001$). Furthermore, for both NH and HI listeners the filter is generally shallower on the low-frequency side

TABLE III. Auditory-filter parameters at 0.8, 1.6, and 3.2 kHz for the normal-hearing listeners. The table also shows, for each listener, the level of the masker spectrum in dB/Hz for each center frequency, and the residual sum of squares of the fitting procedure, ssq. The parameters given are (1) the lower and upper slope of the filter skirt, p_l and p_u, respectively; (2) the dynamic range of the filter, r; and (3) the equivalent rectangular bandwidth of the filter, ERB. The ERB is expressed as a proportion of the center frequency. The value of r is expressed in dB.

subj	f_c kHz	N_0 dB/Hz	p_l	p_u	r	ERB	ssq
1	0.8	52	22.1	28.4	-75.2	0.161	21.1
	1.6	46	21.5	34.6	-72.8	0.151	63.1
	3.2	40	25.0	29.3	-56.5	0.148	122.8
2	0.8	52	16.9	21.6	-82.7	0.211	23.8
	1.6	46	21.3	25.1	-88.3	0.174	24.7
	3.2	40	23.2	28.0	-65.4	0.158	61.5
3	0.8	52	21.7	25.7	-61.6	0.170	68.1
	1.6	46	18.9	29.3	-90.3	0.174	139.2
	3.2	40	23.4	40.5	-52.2	0.135	139.9
4	0.8	52	19.3	25.0	-66.1	0.184	16.0
	1.6	46	23.2	30.8	-69.5	0.151	105.4
	3.2	40	29.0	41.0	-64.8	0.118	22.4
5	0.8	52	20.7	26.7	-85.8	0.172	39.5
	1.6	46	22.4	33.3	-65.4	0.149	8.8
	3.2	40	26.5	37.6	-60.4	0.129	26.4
6	0.8	52	23.0	29.8	-61.4	0.154	23.8
	1.6	46	26.1	36.5	-69.5	0.131	68.7
	3.2	40	23.8	30.8	-55.4	0.149	64.9
7	0.8	52	18.5	24.3	-83.5	0.191	50.7
	1.6	46	20.7	35.0	-109.3	0.154	61.1
	3.2	40	18.8	24.1	-61.1	0.189	36.0
8	0.8	52	18.5	32.6	-74.3	0.170	187.1
	1.6	46	28.5	33.0	-74.3	0.131	45.0
	3.2	40	37.2	28.8	-54.2	0.123	59.1

TABLE IV. As table III, but showing the auditory-filter parameters for hearing-impaired listeners (1-7).

subj	f_c kHz	N_0 dB/Hz	p_l	p_u	r	ERB	ssq
1	0.8	52	15.9	23.8	-49.7	0.210	26.8
	1.6	46	13.6	22.5	-31.7	0.236	16.7
	3.2	40	21.3	26.5	-28.7	0.169	38.9
2	0.8	62	9.1	10.4	-16.4	0.412	21.8
	1.6	56	10.0	16.8	-19.5	0.320	4.2
	3.2	50	9.7	16.8	-27.3	0.325	4.6
3	0.8	57	12.2	16.3	-24.5	0.287	2.3
	1.6	51	48.2	9.8	-22.2	0.247	20.4
	3.2	45	7.5	25.0	-16.8	0.345	51.9
4	0.8	52	10.1	12.3	-29.8	0.362	2.7
	1.6	46	16.9	30.9	-37.0	0.183	5.6
	3.2	40	12.3	20.6	-29.8	0.260	15.1
5	0.8	52	19.1	28.7	-38.7	0.174	17.1
	1.6	46	21.4	28.7	-36.0	0.163	23.3
	3.2	40	12.6	25.9	-19.9	0.236	7.9
6	0.8	52	10.4	16.2	-37.7	0.315	8.2
	1.6	46	7.5	15.0	-46.7	0.399	20.6
	3.2	40	11.8	12.6	-27.0	0.328	7.6
7	0.8	62	13.5	19.8	-35.3	0.249	18.7
	1.6	56	6.8	10.7	-42.1	0.482	82.7
	3.2	50	46.7	7.3	-15.7	0.318	47.6

TABLE V. As Table III, but showing the auditory-filter parameters for hearing-impaired listeners (8-14).

subj	f_s kHz	N_0 dB/Hz	p_l	p_u	r	ERB	ssq
8	0.8	62	13.4	25.9	-59.7	0.226	24.9
	1.6	56	5.6	49.0	-41.6	0.395	56.4
	3.2	50	11.0	18.7	-82.3	0.289	86.9
9	0.8	62	18.1	17.0	-24.0	0.228	15.9
	1.6	56	25.2	34.0	-25.1	0.138	3.4
	3.2	50	14.3	17.9	-22.9	0.252	10.0
10	0.8	62	12.6	17.3	-54.0	0.274	55.3
	1.6	56	14.7	20.6	-31.5	0.233	19.5
	3.2	50	14.4	21.5	-29.9	0.232	30.9
11	0.8	62	15.0	16.9	-35.9	0.252	18.5
	1.6	56	8.9	15.0	-45.7	0.358	25.8
	3.2	50	20.3	23.8	-22.2	0.183	52.7
12	0.8	52	10.3	12.9	-31.7	0.350	10.5
	1.6	46	17.8	39.0	-41.4	0.164	10.6
	3.2	40	14.3	18.5	-26.3	0.248	25.8
13	0.8	62	10.1	43.2	-34.5	0.245	3.9
	1.6	56	8.4	14.7	-29.0	0.374	22.0
	3.2	50	10.5	16.5	-19.6	0.312	10.7
14	0.8	62	14.8	8.3	-39.3	0.378	37.1
	1.6	56	6.1	17.2	-21.8	0.446	5.8
	3.2	50	35.2	11.7	-23.8	0.228	21.4

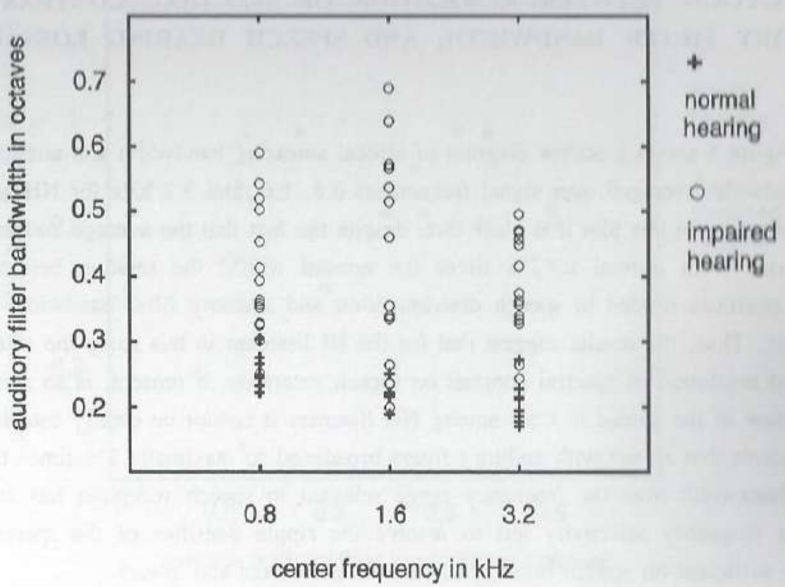


FIG. 4. Auditory filter ERB in octaves for eight normal-hearing and fourteen hearing-impaired listeners at signal frequencies 0.8, 1.6, and 3.2 kHz.

than on the high-frequency side (paired t-test, $p < 0.0001$), with the filters of the NH listeners being more asymmetrical.

Figure 4 shows for NH and HI listeners the ERB in oct at the three signal frequencies. The ERB is quite consistent across NH listeners: on average close to 1/4 oct. This is in good agreement with the results from other studies with normal-hearing listeners (e.g., Dubno and Schaefer, 1992; Moore and Glasberg, 1983; Moore *et al.*, 1990). On the other hand, ERB varies widely among HI listeners, especially at signal frequency 1.6 kHz: from near-normal to about 3 times the normal width. A similar variation in ERB across listeners with mild-to-moderate sensorineural losses was found by Glasberg and Moore (1986), Peters and Moore (1992), and Tyler (1986).

IV. RELATION BETWEEN RESOLUTION OF SPECTRAL CONTRAST, AUDITORY FILTER BANDWIDTH, AND SPEECH HEARING LOSS IN NOISE

Figure 5 shows a scatter diagram of critical smearing bandwidth and auditory filter bandwidth averaged over signal frequencies 0.8, 1.6, and 3.2 kHz for NH and HI listeners. From this plot it is clear that, despite the fact that the average auditory ERB varies from normal to 2½ times the normal width, the relation between spectral contrasts needed in speech discrimination and auditory filter bandwidth is very weak. Thus, the results suggest that for the HI listeners in this study the effect of limited resolution of spectral contrast on speech reception, if present, is so small that in view of the spread in CSB among NH listeners it cannot be clearly established. It seems that an ear with auditory filters broadened to maximally 2½ times the normal bandwidth over the frequency range relevant to speech reception has still sufficient frequency selectivity left to resolve the ripple densities of the spectral envelope sufficient for speech intelligibility (1-1.5 period/oct and lower).

The present findings, indicating that limited resolution of spectral contrast is only loosely associated with hearing loss for speech in noise (see also, experiment I), are consistent with several studies on vowel and consonant identification by sensorineurally hearing-impaired listeners. It is often found that hearing-impaired listeners make few errors in vowel identification (e.g., Owens *et al.*, 1968; Dorman *et al.*, 1985; Hood, 1990), although vowel identification is thought to depend primarily on spectral cues. Thus, even when frequency resolution is reduced, the spectral shapes associated with different vowels are probably preserved. This corroborates the evidence that suggests that accurate vowel recognition requires only resolution of the global spectral shape, rather than resolution of details across the spectrum (e.g., Van Tasell et al., 1987; chapter 2). With respect to consonant identification: although place of articulation depends largely on spectral cues, several researchers have demonstrated that there is no consistent difference between listeners with reduced frequency selectivity and noise masked normals, when equal speech-spectrum audibility is assured across listeners (e.g., Zurek and Delhorne, 1987; Dubno and Schaefer, 1992). Moreover, there are indications that, if differences in consonant identification exist between normal-hearing and hearing-impaired listeners, these are probably more related to deficits in temporal resolution (cf. Rosen and Fourcin, 1986). Obviously, if a listener shows total or near-total degradation in

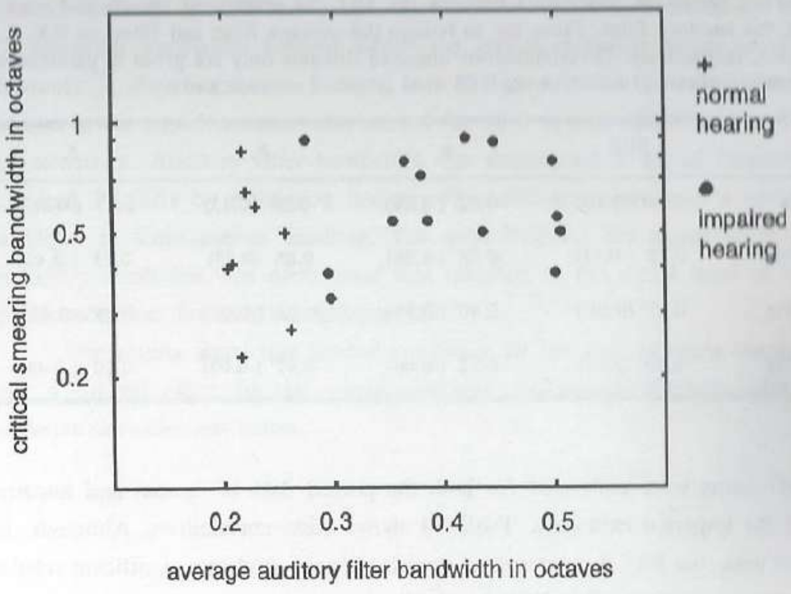


FIG. 5. Critical smearing bandwidth for eight normal-hearing and fourteen hearing-impaired listeners plotted against their auditory filter ERB averaged over 0.8, 1.6, and 3.2 kHz.

frequency resolution, such as HI listener 15, this has a serious effect on speech reception in noise.

Although listeners with a relatively flat loss have been used in the present study, it seems likely that the results apply to other audiometric configurations as well. With respect to the effect of limited resolution of spectral contrasts on speech reception, uniform losses may be considered to represent a worst-case situation.

Together, these results may explain why several attempts to improve speech intelligibility with spectral enhancement for listeners with mild-to-moderate sensorineural losses have been rather disappointing (e.g., Stone and Moore, 1992; Summerfield et al., 1985). It seems that a beneficial effect of enhancing essential spectral contrasts in speech may only be expected in listeners with a rather severe loss of frequency resolution.

In order to assess the general effect of limited frequency resolution on speech hearing loss in noise, we examined the relation between the SRT for unsmeared speech and various auditory-filter characteristics. Pearson product-moment correl-

TABLE VI. Correlation coefficients between the SRT for unsmeared speech and various aspects of the auditory filter. From top to bottom the average filter and filters at 0.8, 1.6, and 3.2 kHz, respectively. Correlations for impaired listeners only are given in parentheses. Correlations significant (1-tailed) at the 0.05 level or better are indicated by *.

	ERB	p_l	p_u	r
mean	0.60* (0.40)	-0.52* (-0.38)	-0.29 (0.15)	0.33 (-0.56*)
0.8 kHz	0.28 (-0.11)	-0.56* (-0.36)	0.05 (0.35)	0.33 (-0.40)
1.6 kHz	0.67* (0.58*)	-0.40* (-0.31)	-0.22 (-0.04)	0.39* (-0.48*)
3.2 kHz	0.50* (0.18)	-0.22 (-0.06)	-0.47* (-0.03)	0.20 (-0.46)

ation coefficients were computed for both the pooled data of normal and impaired ears, and for impaired ears only. Table VI shows these correlations. Although, for the pooled data, the SRT for unsmeared speech shows a moderate significant relation with the average auditory filter ERB (r=0.60), the relation is much weaker and non-significant for HI listeners only. A similar result is obtained for the relation between SRT and the average low-frequency slope of the auditory filter, p_l. This suggests that frequency resolution *per se* is at most one of more parameters associated with speech intelligibility in noise for listeners with flat losses. Similar results were found by Glasberg and Moore (1992), who investigated the relation between several psychoacoustic abilities, including frequency selectivity, and speech reception in listeners with relatively flat cochlear hearing impairments.

Examination of the correlations for the individual center frequencies at which auditory filter characteristics were estimated, revealed similar relations, with the exception that the relation between speech hearing loss in noise and the auditory filter ERB at 1.6 kHz was also significant for the subgroup of HI listeners (r=0.58, p<0.02).

SUMMARY AND CONCLUSIONS

In the present study, we examined for hearing-impaired listeners with mild-to-moderate sensorineural losses the relation between the resolution of spectral contrasts needed in speech discrimination and speech hearing loss in noise. The

spectral contrasts needed for speech discrimination were determined by measuring the smearing bandwidth beyond which the speech-reception threshold (SRT) for sentences in noise increases. Hearing loss for speech in noise was determined by measuring the signal-to-noise ratio needed for 50% correct identification of unsmeared sentences. Auditory filter bandwidth was determined at signal frequencies 0.8, 1.6, and 3.2 kHz by estimating auditory-filter characteristics with a notched-noise paradigm in simultaneous masking. For each listener, the signal level at which frequency resolution was determined was matched to the signal level at which the speech-reception threshold was determined.

The results show that limited resolution of spectral contrasts seems to have only a limited effect on the speech-reception abilities of listeners with mild-to-moderate sensorineural losses.

SUMMARY

The perception of speech is a complex process which, to a certain extent, depends on the peripheral auditory system's ability to resolve spectral features in the acoustic signal. This suggests that a reduction in the ear's ability to resolve spectral features, or a corresponding reduction applied to the signal's spectral features, will have a detrimental effect on speech intelligibility. With respect to spectral features, probably the most salient effect of limited frequency resolution is a blurring of the spectral envelope. Wider auditory filters will preserve less detail of the spectral envelope, thus making it more difficult to distinguish between the many spectrally based speech contrasts. The aim of the present study is to quantify the importance of spectral contrasts to the intelligibility of speech in a critical listening condition.

The signal processing applied in this study to reduce spectral contrasts is based on an analysis-resynthesis system that consists of short-term fast Fourier transforms, spectral modification, inverse transforms, and overlapping additions to reconstruct a continuous signal. Spectral contrasts are reduced by smearing the short-term power-spectrum envelope by a convolution with a Gaussian-shaped filter on a log-frequency scale. In this way spectral contrasts can be gradually reduced, whereas the signal's phase and harmonic structure are preserved. The signal processing may be interpreted as the effect of limited resolution of the power-spectrum envelope which is isolated from auditory processing. The effect of reduced spectral contrast on the intelligibility of speech is mainly assessed by measuring the masked speech-reception threshold (SRT) for sentences. Both the sentences and the masker are subjected to spectral envelope smearing.

After a general introduction in chapter 1, chapters 2 and 3 present a series of listening experiments conducted with, in total, forty normal-hearing listeners. In these experiments the degree of spectral envelope smearing is systematically varied. The first goal of this series was to determine the effect of reduced spectral contrast on sentence intelligibility in a critical listening situation. The SRT in noise for sentences pronounced by a male and a female speaker was measured as a function of the degree of spectral envelope smearing, which varied from smearing bandwidths of 1/8 to 4 octaves. First of all, the results show that spectral envelope smearing over bandwidths up to 1/3 oct does not affect the intelligibility of speech. However, for smearing bandwidths wider than 1/3 oct sentence intelligibility decreases progressively, that is, the masked SRT increases relative to the SRT for unsmeared speech. Secondly, despite the differences between male and female speech reflected in the amount of spectral contrast in the speech signal, including position, amplitude, and bandwidth of formant peaks, and the depth of the valleys in between, the intelligibility of male and female speech is not differentially affected by spectral envelope smearing. The masked SRT increases similarly for both voices for smearing bandwidths over 1/3 oct.

The second goal was to determine how resistant individual vowels and consonants are to a reduction in spectral contrasts. As might be expected, identification of isolated vowels appears to be much more susceptible to severe spectral envelope smearing than identificaton of initial and final consonants in a consonant-vowel-consonant context. The pattern of confusions in a severe condition (2-oct smearing bandwidth) of spectral smearing shows that a predominant proportion of vowels is identified as the vowels /ɔ/ and /u/ (as in the Dutch words "bot" and "voet", respectively). Generally, the target vowels are confused with those particular back vowels that share a first formant position with the target vowel. The results indicate that correct identification is warranted as long as the ripples in the spectral envelope denoting the first and second formant are preserved. Consonants are primarily confused where place of articulation is concerned. Thus, in productive terms, target plosives, fricatives, nasals, and semivowels are predominantly confused with consonants within the same categories. The degree to which each individual target consonant suffers from spectral smearing depends both on the degree to which the target consonant is characterized spectrally and on its distribution of energy over the spectrum.

Finally, the contribution of spectral contrasts to the rather large release from masking found for speech maskers relative to noise maskers is determined. The results show that the extent of the threshold difference between a speech masker and a noise masker decreases as a function of the degree of spectral envelope smearing. This suggests that spectral contrasts contribute to the release from masking mainly because spectral dips in the competing speech facilitate the detection of the target speech by virtue of the favorable signal-to-noise masker ratio in these regions.

An alternative way of quantifying spectral envelope smearing, is in terms of the ripple density (in periods per oct) present in the signal's short-term spectral

envelope before and after smearing. Expressed in these terms, the results from chapters 2 and 3 indicate that the lower spectral ripple densities up to a limit of about 1.5 periods/oct are sufficient for the intelligibility of speech (including vowel and consonant identification) in a critical listening situation. The importance of the corresponding degree of auditory frequency resolution can be illustrated with the following example. Having to understand speech in the presence of a competing speaker at a signal-to-competing-speech ratio of 0 dB is a very common situation, and, normally, a listener is perfectly able to do this. However, if the listener would have a degree of auditory frequency resolution corresponding to, for example, a smearing bandwidth of 2 oct, understanding speech in this situation would be impossible (cf. chapter 3, figures 3 and 4).

An interesting question which follows, is whether limited resolution of spectral contrasts can explain (part of) hearing loss for speech in noise. Chapter 4 examines the relation between the resolution of spectral contrasts needed in speech discrimination and hearing loss for speech in noise for fifteen hearing-impaired listeners with relatively flat, mild-to-moderate sensorineural losses and eight normal-hearing listeners. Auditory filter bandwidth, as a measure of frequency resolution, was determined by estimating auditory filter shapes at center frequencies of 0.8, 1.6, and 3.2 kHz, using a notched-noise masking paradigm. The spectral contrasts essential to speech discrimination were determined by measuring the reduction of spectral contrasts beyond which the speech-reception threshold for sentences in noise increases. Although the average auditory filter bandwidth of the hearing-impaired listeners varies from normal to 2½ times the normal width and the degree of hearing loss for speech in noise varies widely, limited resolution of spectral contrasts seems to be only loosely associated with hearing loss for speech in noise.

In conclusion, the results from this study show that there is a clear relation between spectral contrasts and the intelligibility of speech in a critical listening situation. With respect to hearing loss for speech in noise, the results suggest that limited resolution of spectral contrast is a factor of minor importance.

SAMENVATTING

De auditieve verwerking van spraak is een complex proces waarin onder meer het frequentie-oplossend vermogen van het oor een rol speelt. Dit suggereert een relatie tussen de auditieve frequentieresolutie en spraakverstaan. Omdat spraak in stilte zeer redundant is, wordt het grote nut van de auditieve frequentieresolutie gezocht in ons vermogen onder moeilijke luisteromstandigheden nog spraak te verstaan. Hoewel in de afgelopen twintig jaar het frequentie-oplossend vermogen van het auditieve systeem uitvoerig is onderzocht, en er vaak gewezen is op de relatie met het verstaan van spraak, is de vraag in hoeverre de verstaanbaarheid van spraak ook daadwerkelijk afhangt van het frequentie-oplossend vermogen nog vrijwel niet onderzocht. De moeilijkheid is dat de frequentieselectiviteit van het oor niet rechtstreeks kan worden gevarieerd.

Het frequentie-oplossend vermogen van het perifere auditieve systeem wordt meestal voorgesteld als een serie overlappende filters. Het signaal aan de uitgang van zo'n auditief filter met een vaste centrale frequentie wordt bepaald door de vorm van het filter toegepast op het energiespectrum van het binnenkomende signaal. Verbreding van de auditieve filters leidt dus tot een verminderd frequentie-oplossend vermogen. Dit zou op grofweg twee manieren het verstaan van spraak kunnen beïnvloeden. Ten eerste geven bredere auditieve filters de spectrale eigenschappen van spraak slechter door, waardoor het moeilijker wordt spraakklanken op basis van spectrale eigenschappen te onderscheiden. Wat betreft deze spectrale eigenschappen is het vervagen van de spectrale omhullende waarschijnlijk de belangrijkste factor. Ten tweede zullen bredere auditieve filters in een situatie waarbij spraak moet worden verstaan tegen een achtergrond van stoorlawaai meer lawaai doorlaten, waardoor het verstaan moeilijker wordt. Een verminderd frequentie-oplossend vermogen is dan ook regelmatig in verband gebracht met gehoorverlies voor spraakverstaan in ruis, zoals kan optreden bij slechthorendheid.

Een methode om zonder de frequentieresolutie van het oor te veranderen toch iets te weten te komen over het belang van deze eigenschap voor het verstaan van spraak, is de spectrale informatie in het *spraaksignaal* dat aan het oor wordt aangeboden te degraderen en het effect van deze degradatie op de verstaanbaarheid te meten. In deze aanpak worden dus uit het oogpunt van auditieve frequentieresolutie die aspecten van het spraaksignaal gevarieerd die belangrijk geacht worden voor het verstaan van spraak.

Het doel van dit onderzoek is het belang van spectrale contrasten voor de verstaanbaarheid van spraak te bepalen. Afgezien van het wetenschappelijke belang een beter inzicht te krijgen in de spectrale eigenschappen van spraak en de rol van auditieve frequentieresolutie in spraakverstaan, kunnen de gegevens ook gebruikt worden voor meer praktische toepassingen, zoals de codering en transmissie van spraaksignalen, en de signaalbewerking ter verbetering van de verstaanbaarheid van spraak voor slechthorenden.

De in deze studie toegepaste signaalbewerking om spectrale contrasten te verminderen is gebaseerd op een analyse- en resynthesesysteem dat bestaat uit Fourier transformaties op korte segmenten van het spraaksignaal en na bewerking een overlappende optelling van die segmenten om een continu signaal te reconstrueren. De spectrale contrasten worden verminderd door een convolutie van de omhullende van het energiespectrum met een Gaussisch filter op een logaritmische frequentie-as. Op deze manier kunnen de spectrale contrasten geleidelijk worden verminderd, terwijl de harmonische structuur en de daarbij behorende faserelaties in het spraaksignaal bewaard blijven.

Als maat voor de verstaanbaarheid van spraak is de drempel voor het verstaan van eenvoudige zinnen tegen een achtergrond van stoorlawaai, de "speech-reception threshold" of SRT, genomen. De SRT is gedefinieerd als die signaal-ruisverhouding in decibels waarbij een luisteraar 50% van de zinnen zonder fouten kan reproduceren. Spectrale contrasten werden zowel in de zinnen als in het stoorsignaal verminderd.

Na een algemene inleiding in hoofdstuk 1, wordt in de hoofdstukken 2 en 3 een serie luisterproeven met in totaal veertig normaalhorende luisteraars beschreven. In deze experimenten wordt de mate van spectrale versmering systematisch gevarieerd. Het primaire doel van deze serie experimenten is het effect van verminderde spectrale contrasten op de verstaanbaarheid van zinnen onder moeilijke luisteromstandigheden te bepalen. Hiertoe wordt de SRT in ruis gemeten als functie van de mate van spectrale versmering, zowel voor zinnen uitgesproken door een vrouwenstem als voor zinnen uitgesproken door een mannenstem. De bandbreedte van de versmering varieert van 1/8 tot 4 octaven. De resultaten laten zien dat versmering van de spectrale omhullende over een bandbreedte tot 1/3 octaaf geen invloed heeft op het spraakverstaan. Echter, voor versmering over meer dan 1/3 octaaf neemt de zinsverstaanbaarheid progressief af, dat wil zeggen, de SRT in ruis neemt toe ten opzichte van de SRT voor onversmeerde spraak. Verder laten de resultaten zien dat

de verstaanbaarheid van mannen- en vrouwenspraak niet verschillend te lijden heeft onder versmering van de spectrale omhullende, ondanks de verschillen tussen mannen- en vrouwenspraak in spectrale contrasten in het spraaksignaal, inclusief de positie, amplitude, en bandbreedte van de formantpieken en de diepte van de dalen daartussen. Voor beide stemmen neemt de SRT in ruis op dezelfde wijze toe voor versmering over meer dan 1/3 octaaf.

Een tweede doel van deze experimenten is te bepalen hoe goed individuele klinkers en medeklinkers bestand zijn tegen een vermindering van spectraal contrast. Zoals valt te verwachten, blijkt de identificatie van losse klinkers veel gevoeliger te zijn voor ernstige versmering van de spectrale omhullende dan de identificatie van begin- en eindmedeklinkers in betekenisloze woordjes van het type medeklinker-klinker-medeklinker. Uit de verwarringsmatrix verkregen bij sterke versmering van de spectrale omhullende (versmering over 2 octaven) blijkt dat een groot aantal klinkers verward wordt met de klinkers /ɔ/ en /u/ (als in de woorden "bot" en "voet"). Over het algemeen worden de klinkers verward met een achterklinker waarvan de eerste formant vergelijkbaar is met die van de te identificeren klinker. De resultaten suggeren dat een correcte identificatie van klinkers gewaarborgd is zolang de modulaties die de eerste en tweede formant aangeven in de spectrale omhullende behouden blijven. Medeklinkers worden vooral verward wat betreft de plaats van articulatie. Dus, gesteld in productietermen, plofklanken, wrijfklanken, nasalen, en halfklinkers worden voornamelijk verward met medeklinkers binnen dezelfde klasse. Verder hangt de mate waarin een medeklinker gevoelig is voor spectrale versmering zowel af van de mate waarin die medeklinker spectraal bepaald is als van de manier waarop de energie van die medeklinker verspreid is over het spectrum.

Tenslotte wordt de bijdrage van spectrale contrasten aan het nogal grote verschil in maskering tussen spraak en ruis als stoorder onderzocht (spraak wordt makkelijker verstaan tegen een achtergrond van stoorspraak dan stoorruis). De resultaten laten zien dat dit verschil afneemt als functie van de mate van versmering van de spectrale omhullende. Dit wijst er op dat spectrale contrasten bijdragen aan het verschil in maskering voornamelijk omdat de spectrale dalen in de stoorspraak de te volgen spraak helpen detecteren door hun gunstige signaal-ruisverhouding.

Een alternatieve manier om het effect van versmering van de spectrale omhullende uit te drukken is in termen van de dichtheid van de modulaties (in perioden per octaaf) in de spectrale omhullende van het signaal voor en na bewerking. Hierin uitgedrukt, geven de resultaten uit de hoofdstukken 2 en 3 aan dat spectrale modulaties tot ongeveer 1.5 periode/octaaf al voldoende zijn voor de verstaanbaarheid van spraak onder moeilijke luisteromstandigheden (inclusief identificatie van klinkers en medeklinkers). Een voorbeeld kan het belang van een hiermee corresponderende auditieve frequentieresolutie duidelijk maken. Het verstaan van spraak in aanwezigheid van een storende spreker, met een signaal-stoorverhouding van 0 dB, is een alledaagse situatie, waarmee een goedhorende luisteraar geen enkel probleem heeft. Echter, als de mate van auditieve frequentieresolutie van deze luisteraar zou overeenkomen met een filter van 2 octaven, dan zou deze luisteraar in de gegeven situatie onmogelijk de bedoelde spraak kunnen verstaan (vgl. hoofdstuk 3, figuren 3 en 4).

Een interessante vraag is of een verminderd vermogen om spectrale contrasten op te lossen het gehoorverlies voor spraakverstaan in ruis zoals dat voorkomt bij slechthorenheid zou kunnen verklaren. Hoofdstuk 4 onderzoekt deze vraag bij vijftien luisteraars met een beperkt, vlak perceptief gehoorverlies en acht normaalhorende luisteraars. Als maat voor het auditief frequentie-oplossend vermogen, wordt de vorm en de breedte van het auditieve filter gemeten met behulp van een "notched-noise" maskeerparadigma bij de middenfrequenties 0.8, 1.6, and 3.2 kHz. De spectrale contrasten die nodig zijn voor het spraakverstaan worden bepaald door de spectrale contrasten in spraak te verminderen en die versmering te bepalen waarbij de verstaanbaarheid begint af te nemen. Hoewel de gemiddelde bandbreedte van het auditieve filter bij de slechthorende luisteraars varieert van normaal tot 2.5 keer breder dan normaal, is de relatie tussen de resolutie van spectrale contrasten nodig voor spraakverstaan en het gehoorverlies voor spraakverstaan in ruis zwak.

Samenvattend, tonen de resultaten van dit proefschrift aan dat er een duidelijke relatie is tussen spectrale contrasten en de verstaanbaarheid van spraak onder moeilijke luisteromstandigheden. Wat betreft gehoorverlies voor spraakverstaan in ruis lijkt een verminderd vermogen om spectrale contrasten op te lossen een factor van ondergeschikte betekenis te zijn.

REFERENCES

Allen, J.B., and Rabiner, L.R. (1977). "A unified approach to short-time Fourier analysis and synthesis," IEEE 65 (11), 1558-1564.

Beddor, P.S., and Hawkins, S. (1990). "The influence of spectral prominence on perceived vowel quality," J. Acoust. Soc. Am. 87, 2684-2704.

Bonding, P. (1979). "Frequency selectivity and speech discrimination in sensorineural hearing loss," Scand. Audiol. 8, 205-215.

Bosman, A.J. (1989). "Speech perception by the hearing impaired," Ph.D. thesis, Utrecht, The Netherlands.

Carhart R., and Tillman T.W. (1970). "Interaction of competing speech signals with hearing losses," Arch. Otolaryng. 91, 273-279.

Celmer, R.D., and Bienvenue, G.R. (1987). "Critical bands in the perception of speech signals by normal and sensorineural hearing loss listeners," in *The psychophysics of speech perception* edited by M.E.H. Schouten (NATO ASI Series, 1987), Series D, Behavioural and Social Sciences 39, pp. 473-480.

Childers, D.G., and Wu, K. (1991). "Gender recognition from speech. Part II: Fine analysis," J. Acoust. Soc. Am. 90, 1841-1856.

Delattre, P., Liberman, A.M., Cooper, F.S., and Gerstman, L.J. (1952). "An experimental study on the acoustic determinants of vowel color; Observations on one- and two-formant vowels synthesized from spectrographic patterns," WORD 8 (3), 195-210.

Dixon, W.J. (1985). *BMDP Statistical software* (University of California Press, Berkeley).

Dorman, M.F., Marton, K., and Hannley, M.T. (1985). "Phonetic identification by elderly normal and hearing-impaired listeners," J. Acoust. Soc. Am. 77, 664-670.

Dreschler, W.A., and Plomp, R. (1985). "Relations between psychophysical data and speech perception for hearing-impaired subjects. II," J. Acoust. Soc. Am. 78, 1261-1270.

Dubno, J.R., and Dorman, M.F. (1987). "Effects of spectral flattening on vowel identification," J. Acoust. Soc. Am. 82, 1503-1511.

Dubno, J.R., and Schaefer, A.B. (1989). "Frequency resolution for broadband-noise masked normal listeners," J. Acoust. Soc. Am. Suppl. 1 **86**, S122.

Dubno, J.R., and Schaefer, A.B. **(1992)**. "Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners," J. Acoust. Soc. Am. **91**, 2110-2121.

Duquesnoy, A.J. **(1983)**. "Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons," J. Acoust. Soc. Am. **74**, 739-743.

Fant, G. **(1960)**. *Acoustic theory of speech production* (Mouton, 's-Gravenhage).

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

Festen, J.M., and Plomp, R. **(1990)**. "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725-1736.

Fujimura, O., and Lindqvist, J. **(1971)**. "Sweep-tone measurements of vocal-tract characteristics," J. Acoust. Soc. Am. **49**, 541-558.

Gagné, J.P., and Erber, N.P. **(1987)**. "Simulation of sensorineural hearing impairment," Ear Hear. **8**, 232-243.

Glasberg, B.R., and Moore, B.C.J. **(1986)**. "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020-1033.

Glasberg, B.R., and Moore, B.C.J. **(1990)**. "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103-138.

Glasberg, B.R., and Moore, B.C.J. **(1992)**. "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech," Scand. Audiol. Suppl. **32**, 3-25.

Hays, W.L. **(1988)**. *Statistics* (Holt, Rinehart, and Winston, New York).

Hood, J.D. **(1990)**."The pattern of consonant and vowel errors in patients with conductive and cochlear hearing loss," Br. J. Audiol. **24**, 167-169.

Horst, J.W. **(1982)**. "Discrimination of complex signals in hearing," Ph.D. thesis, Groningen, The Netherlands.

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.

Howard-Jones, P.A., Summers, I.R., Al-Dabbagh, A.D., and Garnham, C.W. **(1991)**. "Perception of spectrally degraded speech," personal communication.

Houtgast, T. **(1974)**. "Lateral suppression in hearing," Ph.D. thesis, Amsterdam, The Netherlands.

Klein, W., Plomp, R., and Pols, L.C.W. **(1970)**. "Vowel spectra, vowel spaces, and vowel identification," J. Acoust. Soc. Am. **48**, 999-1009.

ter Keurs, M., Festen, J.M., and Plomp, R. **(1989a)**. "Effects of spectral envelope smearing on speech reception", J. Acoust. Soc. Am. Suppl. 1 **86**, S80.

ter Keurs, M., Plomp, R., and Festen, J.M. **(1989b)**. "Effects of spectral smearing on speech perception," Proc. Europ. Conf. Speech Commun. Technol. **1**, 251-253.

ter Keurs, M., Festen, J.M., and Plomp, R. **(1991)**. "Effect of spectral envelope smearing on speech reception in noise and competing speech," J. Acoust. Soc. Am. **89**, 1960.

ter Keurs, M., Festen, J.M., and Plomp, R. **(1992a)**. "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. **(1992b)**. "Effect of spectral envelope smearing on speech reception. II," accepted for publication in J. Acoust. Soc. Am.

ter Keurs, M., Festen, J.M., and Plomp, R. **(1992c)**. "Limited resolution of spectral contrast and hearing loss for speech in noise," submitted for publication in J. Acoust. Soc. Am.

ter Keurs, M., Festen, J.M., and Plomp, R. **(1992d)**. "Effect of spectral envelope smearing on phoneme identification," in *The psychophysics of speech perception II* edited by M.E.H. Schouten (*in press*).

Leshowitz, B. **(1977)**. "Speech intelligibility in noise for listeners with sensorineural hearing damage," IPO Ann. Prog. Rep. **12**, Eindhoven, The Netherlands.

Levitt, H. **(1971)**. "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467-477.

Lutfi, R.A., and Patterson, R.D. (1984). "On the growth of masking asymmetry with stimulus intensity," J. Acoust. Soc. Am 76, 739-745.

Miller, G.A., and Nicely, P.E. (1954). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. 27, 338-352.

Moore, B.C.J. (1986). *Frequency selectivity in hearing* (AP, London).

Moore, B.C.J., and Glasberg, B.R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. 74, 750-753.

Moore, B.C.J., Peters, R.W., and Glasberg, B.R. (1990). "Auditory filter shapes at low center frequencies," J. Acoust. Soc. Am. 88, 132-140.

Neter, J., Wasserman, W. and Kutner, M.H. (1990). *Applied linear statistical models* (Irwin, Homewood), 3rd ed.

Nierop van, D.J.P.J., Pols, L.C.W., and Plomp, R. (1973). "Frequency analysis of Dutch vowels from 25 female speakers," Acustica 29, 110-118.

Nooteboom, S.G., and Cohen A. (1984). *Spreken en Verstaan* (Van Gorcum, Assen), 2nd ed., Chap. 3.

Owens, E., Talbott, C., and Schubert, E. (1968). "Vowel discrimination of hearing-impaired listeners," J. Speech Hear. Res. 11, 648-655.

Patterson, R.D. (1976). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. 59, 640-654.

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

Peters, R.W., and Moore, B.C.J. (1992). "Auditory filter shapes at low center frequencies in young and elderly hearing-impaired subjects," J. Acoust. Soc. Am. 91, 256-266.

Peterson, G.E., and Barney, H.L. (1952). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. 24, 175-184.

Pick, G.F. (1980). "Level dependence of psychophysical frequency resolution and auditory filter shape," J. Acoust. Soc. Am. 68, 1085-1095.

Plomp, R. (1983). "Perception of speech as a modulated signal," Proc. Tenth Intern. Congr. Phonetic Sciences, 29-40.

Plomp, R. (1986). "A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired," J. Speech Hear. Res. 29, 146-154.

Plomp, R., and Mimpen, A.M. (1979). "Improving the reliability of testing the speech-reception threshold for sentences," Audiology 18, 43-52.

Plomp, R., and Steeneken, H.J.M. (1969). "Effect of phase on the timbre of complex tones," J. Acoust. Soc. Am. 46, 409-421.

Pols, L.C.W. (1977). "Spectral analysis and identification of Dutch vowels in monosyllabic words," Ph.D. thesis, Amsterdam, The Netherlands.

Rosen, S., and Fourcin, A. (1986). "The perception of speech," in *Frequency selectivity in hearing* edited by B.C.J. Moore (AP, London), pp. 373-487.

Stelmachowicz, P.G., Jesteadt, W., Gorga, M.P., and Mott, J. (1985). "Speech perception ability and psychophysical tuning curves in hearing-impaired listeners," J. Acoust. Soc. Am. 77, 620-627.

Stone, M.A., and Moore, B.C.J. (1992). "Spectral feature enhancement for people with sensorineural hearing impairment: Effects on speech intelligibility and quality," J. Rehab. Res. & Devel. 29, 39-56.

Summerfield, A.Q., Foster, J., Tyler, R., and Bailey, P.J. (1985). "Influences of formant narrowing and auditory frequency selectivity on identification of place of articulation in stop consonants," Speech Commun. 4, 213-229.

Turner, Ch.W., and Van Tasell, D.J. (1984). "Sensorineural hearing loss and the discrimination of vowel-like stimuli," J. Acoust. Soc. Am. 75, 562-565.

Tyler, R.S. (1986). "Frequency resolution in hearing-impaired listeners," in *Frequency selectivity in hearing*, edited by B.C.J. Moore (AP, London), pp. 309-371.

Van Tasell, D.J., Fabry, D.A., and Thibodeau, L.M. (1987). "Vowel identification and vowel masking patterns of hearing-impaired subjects," J. Acoust. Soc. Am. 81, 1586-1597.

Villchur, E. (1977). "Electronic models to simulate the effect of sensory distortions on speech perception by the deaf," J. Acoust. Soc. Am. 62, 665-674.

van Veen T.M., and Houtgast, T. (1985). "Spectral sharpness and vowel dissimilarity," J. Acoust. Soc. Am. 77, 628-634.

Zurek, P.M., and Delhorne, L.A. (1987). "Consonant reception in noise by listeners with mild to moderate sensorineural hearing impairment," J. Acoust. Soc. Am. 82, 1548-1559.

CURRICULUM VITAE

Mariken ter Keurs werd op 19 juli 1963 geboren te Rijswijk (ZH). In 1981 behaalde zij het diploma Gymnasium-A met wiskunde en biologie aan het Revius Lyceum te Doorn. Vervolgens studeerde zij Klassieke Taal- en Letterkunde aan de Rijksuniversiteit Utrecht. Na het kandidaatsexamen op 5 juli 1984 studeerde zij Experimentele Fonetiek met als bijvakken o.a. Experimentele Audiologie (electrische binnenoorprothese), Informatica en Klassiek Grieks. Het afstudeerwerk betrof een onderzoek naar het verschijnsel van spectrale synchronie in lopende spraak en het werd uitgevoerd bij de afdeling Audiologie van het Instituut voor Zintuigfysiologie te Soesterberg. Op 10 december 1987 behaalde zij het doctoraaldiploma Experimentele Fonetiek (oude stijl).

In de periode van 16 december 1987 tot 15 december 1991 voerde zij, in dienst van de Vrije Universiteit van Amsterdam, het in dit proefschrift beschreven onderzoek uit bij de afdeling Experimentele Audiologie van de vakgroep KNO, faculteit der Geneeskunde.

8. Het algemeen gebruik van bloedtesten in de pluimveehouderij met onbekende sensitiviteit en specificiteit geeft (te) veel ruimte voor subjectiviteit in de beoordeling van de gevonden waarden.

9. Voor de consument met een groot vertrouwen in de woorden natuurproduct, natuurlijk, biologisch e.d. is het goed te weten dat de natuur producten oplevert die giftig en kankerverwekkend zijn, zoals bijvoorbeeld aflatoxine B1 (o.a. voorkomend op pinda's).

10. De geloofwaardigheid en het belang van veterinaire keuringen bij menwedstrijden zou toenemen, wanneer de uitslag van de keuring bindend zou zijn.

11. Je kunt in Nederland anno 1992 beter een zieke koe dan een zieke veehouder zijn, wanneer buiten praktijkuren (lees kantooruren) hulp aan huis van een dierenarts respectievelijk huisarts gewenst is.

12. Doodsoorzaak nummer 1 heeft het eeuwige leven als bron van onderzoek.

INTELLIGIBILITY OF SPECTRALLY SMEARED SPEECH

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan
de Vrije Universiteit te Amsterdam,
op gezag van de rector magnificus
dr. C. Datema,
hoogleraar aan de faculteit der letteren,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der geneeskunde
op vrijdag 27 november 1992 te 10.30 uur
in het hoofdgebouw van de universiteit, De Boelelaan 1105

door

MARIKEN TER KEURS

geboren te Rijswijk (ZH)

Bariet, Ruinen
1992