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

Nicolle H. van Schijndel,[b] Tammo Houtgast, and Joost M. Festen[c]
Department of Otolaryngology, University Hospital VU, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

(Received 8 November 1999; accepted for publication 16 April 2001)

Many hearing-impaired listeners suffer from distorted auditory processing capabilities. This study examines which aspects of auditory coding (i.e., intensity, time, or frequency) are distorted and how this affects speech perception. The distortion-sensitivity model is used: The effect of distorted auditory coding of a speech signal is simulated by an artificial distortion, and the sensitivity of speech intelligibility to this artificial distortion is compared for normal-hearing and hearing-impaired listeners. Stimuli (speech plus noise) are wavelet coded using a complex sinusoidal carrier with a Gaussian envelope ($\frac{1}{4}$ octave bandwidth). Intensity information is distorted by multiplying the modulus of each wavelet coefficient by a random factor. Temporal and spectral information are distorted by randomly shifting the wavelet positions along the temporal or spectral axis, respectively. Measured were (1) detection thresholds for each type of distortion, and (2) speech-reception thresholds for various degrees of distortion. For spectral distortion, hearing-impaired listeners showed increased detection thresholds and were also less sensitive to the distortion with respect to speech perception. For intensity and temporal distortion, this was not observed. Results indicate that a distorted coding of spectral information may be an important factor underlying reduced speech intelligibility for the hearing impaired. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1378345]

PACS numbers: 43.71.Ky, 43.71.An, 43.66.Dc, 43.66.Fe, 43.66.Mk [RVS]

I. INTRODUCTION

The difficulty hearing-impaired listeners have in perceiving speech in noise has been the subject of many investigations, but is still not entirely understood. Although audibility plays an important role, several studies have shown that this cannot explain the whole problem [see, e.g., Moore (1996) or Noordhoek *et al.* (2000)]. These studies have demonstrated that factors apart from reduced audibility, called suprathreshold deficits, degrade speech processing. Suprathreshold deficits can distort the auditory processing of either intensity, time, or frequency information, or a combination of these types of information. For example, excessive forward and backward masking are consequences of suprathreshold deficits that may be reduced to a single factor of distorted temporal coding; excessive upward and downward spread of masking may be related to distorted spectral coding. Impaired loudness perception probably relates to a distorted representation of intensity information. This study evaluates these three types of information. The aim is to investigate how reduced speech intelligibility relates to distorted coding of intensity, time, or frequency.

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

The correlation approach results in statistical relations between reduced speech perception and suprathreshold deficits. A drawback of this approach is that one cannot exclude that an underlying common factor causes the observed correlation. For example, if a correlation between speech intelligibility and spectral resolution is observed, an underlying common factor can be the hearing threshold. Then, higher hearing thresholds instead of reduced frequency selectivity may cause reduced speech perception. In different studies, underlying factors probably had different effects, which may explain the different results. Relations between distorted auditory coding and speech perception can be investigated in a

---

[a] Part of the results was presented at the Joint ASA/EAA/DEGA Meeting ''Berlin 99'' [Collected papers from the joint meeting ''Berlin 99,'' Deutsche Gesellschaft für Akustik, Oldenburg].

[b] Current address: Philips Research Laboratories, Digital Signal Processing (WY 82), Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands; electronic mail: Nicolle.van.Schijndel@philips.com

[c] Electronic mail: JM.Festen@azvu.nl

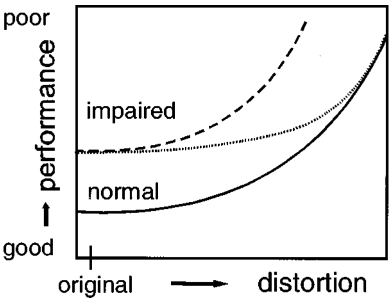

FIG. 1. Illustration of the distortion-sensitivity model. Performance for hearing-impaired listeners as a function of distortion is compared with that of normal-hearing listeners (solid line). The possible outcome of such an experiment is ‘‘convergence’’ (dotted and solid lines) or ‘‘no convergence’’ (dashed and solid lines).

more direct way using the distortion-sensitivity model (Houtgast, 1995; van Schijndel *et al.*, 2001).

Under the distortion-sensitivity model (Fig. 1), the relation between speech intelligibility and distorted *auditory* coding is studied by simulating the effect of the auditory deficit by *artificial* distortion of the speech signal. The idea is that removing cues that are not perceived by the hearing impaired will not affect their performance. Performance is measured as a function of distortion, and compared for normal-hearing and hearing-impaired listeners. Two trends may be observed: *convergence* (dotted and solid lines) or *no convergence* (dashed and solid lines). In the convergence case, hearing-impaired listeners are less sensitive to the distortion than normal-hearing listeners. Then, it may be concluded that the artificial distortion relates to distorted auditory coding that impedes performance. The artificial distortion affects the sound characteristics in the same way as the auditory deficits. In the no-convergence case, hearing-impaired listeners are as sensitive to the distortion as normal-hearing listeners, indicating that the artificial distortion has no relation to the hearing deficits causing difficulties in speech perception. In the no-convergence case, performances of the normal-hearing and hearing-impaired listeners will run parallel, or may even diverge. Divergence will be discussed in more detail in Sec. III C 2. A few studies (Duquesnoy and Plomp, 1980; ter Keurs *et al.*, 1993; Turner *et al.*, 1995; van Schijndel *et al.*, 2001) used the principles of the distortion-sensitivity model so far, but they did not explicitly explain their results in terms of the model, except the last study.

In van Schijndel *et al.* (2001) the distortion-sensitivity model was used with respect to the coding of intensity information. It was concluded that reduced intensity coding accuracy may partly explain impaired speech perception.

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

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

The studies mentioned previously obtained data that can be analyzed in terms of the distortion-sensitivity model. The effects of distortion of intensity, time, and frequency information on speech perception were studied in isolation, although these three domains are not completely independent. Manipulation in one domain will affect the other domains. For example, spectral smearing introduces temporal smearing and vice versa. In Sec. II A 4, this will be illustrated. Awareness of these unwanted by-products of the speech-processing algorithm is important. Therefore, in the present study, the interdependency of the intensity, time, and frequency domains was taken into account.

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

II. METHOD

A. Degradation of intensity, time, and frequency information

In this study, a sound-processing algorithm is used to degrade artificially the intensity, time, and frequency content of speech. The degradation is intended to simulate the effects of distorted auditory coding. By means of the speech-reception threshold test (SRT, Sec. II D 3 a), speech intelligibility of sentences was measured as a function of applied artificial distortion. In order to simulate auditory coding, a perceptually relevant spectro-temporal decomposition and recomposition method was developed. This method was also used in van Schijndel *et al.* (2001), and is described in the following.

1. Spectro-temporal decomposition and recomposition

To model auditory spectro-temporal coding, sounds were described in the time–frequency domain by means of a wavelet transform. Compared with the short-time Fourier transform, the wavelet transform matches auditory system coding more closely because it uses a logarithmic frequency scale (e.g., Rioul and Vetterli, 1991). An important criterion in the choice of the mother wavelet is its spectral and temporal width. Results of van Schijndel *et al.* (1999) suggest that a Gaussian-windowed sinusoid with a shape factor between 0.15 and 0.3 roughly matches the auditory time-frequency window. Therefore, as the prototype analysis function, a Gaussian wavelet was chosen. The Gaussian wavelet is a complex sinusoidal carrier with a Gaussian envelope:

$$s(t)=\sqrt{\alpha f_0}\exp(i2\pi f_0 t)\exp(-\pi(\alpha f_0 t)^2), \tag{1}$$

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

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

For simultaneous sampling in time and frequency, the Nyquist sampling theorem was applied twice (Allen, 1977; Allen and Rabiner, 1977). The sampling interval was based on the temporal and spectral range over which the Gaussian wavelet is essentially different from zero. Since the Gaussian wavelet does not have compact support[1] in time, or in frequency, the range between the points that were 25 dB down from the peak was taken as the range over which the window is significant, i.e., essentially different from zero. Thus, outside these 25-dB down points, the window is considered to be negligible. This definition corresponds to a duration of about twice the effective duration and a bandwidth of about twice the effective bandwidth. Application of this criterion leads to a sampling of one wavelet per three periods of the wavelet carrier frequency along the time axis, and eight wavelets per octave along the frequency axis. Theoretically, the number of complex coefficients needed to describe the signal using the 25-dB criterion for sampling is about two coefficients per input sample (Allen, 1977). In this study, the frequency of the signals was limited to the range from 250 to 4000 Hz. As a result, 1 s of speech (sampling frequency: 44.1 kHz; no information below 250 Hz or above 4 kHz preserved) was described by 16×10^3 complex wavelet coefficients.

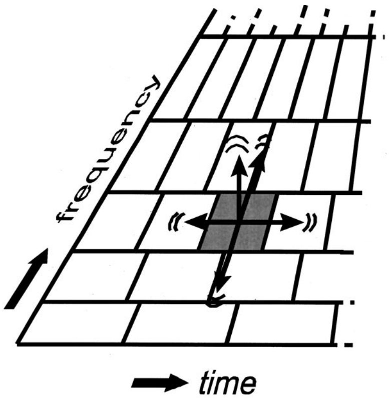

FIG. 2. Schematic illustration of the perturbation of the intensity, time, or spectral information. The Gaussian wavelets are symbolized by rectangles. Each wavelet is given a random perturbation with respect to its intensity, temporal position, or spectral position.

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

Between decomposition and recomposition, the integrity of the intensity, time, or frequency information was reduced to simulate poor auditory coding. Intensity degradation was obtained by introducing uncertainty in the modulus of each wavelet coefficient. Temporal and spectral degradations were obtained by introducing uncertainty in the temporal and spectral position of each wavelet, respectively. In Fig. 2, this is illustrated schematically. In the following paragraphs, these different types of degradation will be explained in more detail. After the perturbation, the energy contained in each frequency band over the whole test sentence was scaled to equal the original energy in that band. Since this study aims at investigating speech-perception performance in noise, speech and noise were summed before processing.

2. Degradation of the intensity accuracy

To degrade the accuracy of the intensity information, the modulus of the wavelet coefficients was perturbed (intensity perturbation). This was achieved by multiplying each wavelet coefficient by a random factor. As a result, silence will remain silence after perturbation. The random perturbation factor ε (in dB) was chosen from a uniform distribution with zero mean and boundaries[2] $-L_D/2$ and $+L_D/2$. Thus the modulus of each individual coefficient was multiplied by a different random factor $10^{\varepsilon/20}$. The intensity perturbation levels used in this study were moderate and the uncomfortable loudness levels of the subjects were never approached.

3. Degradation of the temporal accuracy

To degrade the accuracy of the temporal information, the positions of the wavelets were shifted randomly along the temporal axis (temporal perturbation). To avoid a degradation of the accuracy of spectral information as much as possible, only the temporal envelope of the wavelets was displaced, not the underlying fine structure. The new fine structure was calculated by extrapolation of the original fine structure to the new position of the envelope. As a result, the information contained within the original fine structure was left unaffected. The position of the envelope of each wavelet was shifted independently by a random value chosen from a uniform distribution ranging from $-T_D/2$ to $+T_D/2$. The degree of temporal distortion T_D is expressed in terms of the duration of the wavelets (inversely proportional to the bandwidth). If T_D equals two wavelets, the maximal displacement along the time axis is one effective duration of the wavelet from its original position. At 1 kHz, this is 5.76 ms; at 4 kHz, this is 1.55 ms.

4. Degradation of the spectral accuracy

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

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

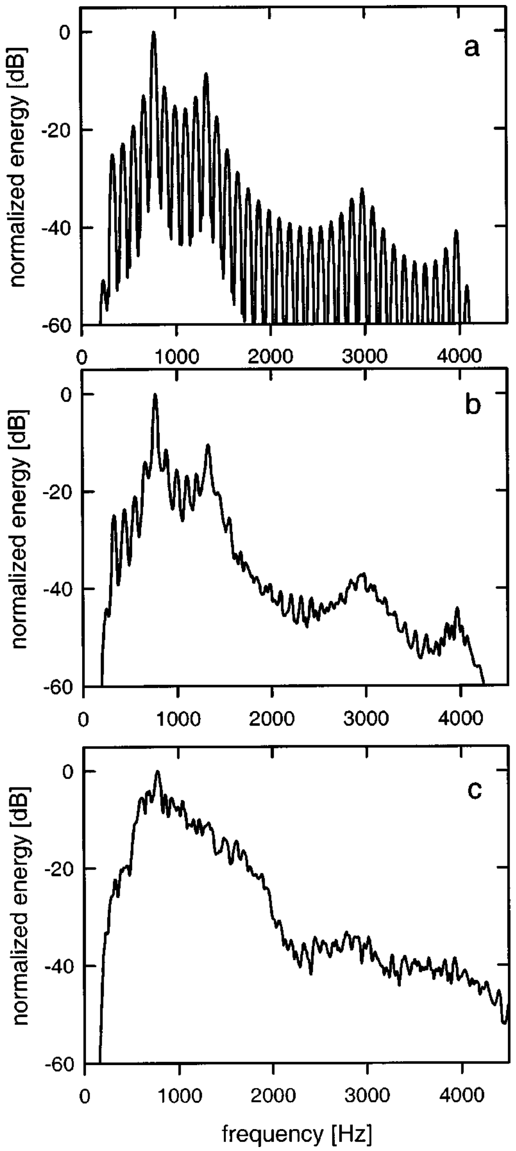

FIG. 3. The effect of the artificial distortion of the spectral information on an artificial vowel/a/: (a) undistorted vowel, (b) spectral reference condition (phase distorted), (c) spectral perturbation of 0.75 octaves (phase distorted and spectrally perturbed).

In Fig. 3, the effect of distorting the spectral information of an artificial vowel /a/ is illustrated. Panel (a) shows the undistorted vowel. In panel (b), the vowel is plotted in the spectral reference condition. In this condition, the phase of the complex coefficients is distorted, but the positions of the wavelets along the spectral axis are retained. As a result, most of the spectral fine structure is lost, but the spectral envelope is intact. In panel (c), the vowel is plotted in the most severe spectral distortion condition used in this study,

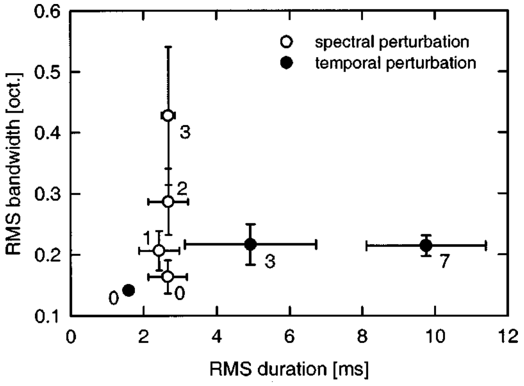

FIG. 4. The effect of the nondeterministic perturbation process on the rms duration (footnote 4) and rms bandwidth of a Gaussian-windowed tone with a center frequency of 1 kHz and a shape factor of 0.1735, i.e., an effective bandwidth of $\frac{1}{4}$ octave. Closed and open symbols represent the values corresponding to temporal perturbation and spectral perturbation, respectively. The numbers represent the degree of perturbation (expressed in the number of wavelets). The error bars represent the standard deviation.

i.e., when F_D equals 0.75 octaves. The phase is distorted as in the reference condition, and in addition the wavelets were shifted randomly over a maximum of $F_D/2$ along the spectral axis. As a result, the spectral envelope is smeared almost fully. Thus the overall spectral effect of the applied spectral uncertainty is a broadening of the spectral peaks.

As mentioned in Sec. I, degradation of the accuracy of the information of one domain is not possible without collateral degradation of the information of other domains. For example, the degradation of the accuracy of the intensity information also affects the spectral and temporal content of a signal. The effects of distortion of temporal information on spectral information and vice versa are illustrated in Fig. 4 for a Gaussian-windowed sinusoid (center frequency $=1$ kHz; $\alpha=0.1735$). This Gaussian tone pulse was wavelet decomposed, followed by spectral or temporal degradation of the wavelet coefficients, and recomposed. For recomposition of the unperturbed signal, the original wavelet coefficients were used. Then, reconstruction is perfect (see Sec. II A 1). The rms duration[4] and rms bandwidth of this signal are indicated by the closed circle with index ''0.'' If the wavelet coefficients are perturbed, the reconstructed signal is different from the original signal. Temporal or spectral perturbation of the coefficients increase the duration and bandwidth of the reconstructed signal. The effects of temporal perturbation on the duration and bandwidth of the signal are represented by the other closed circles; the effects of the spectral perturbation are indicated by open circles. For each degree of distortion, the perturbation procedure was applied to the input signal six times. Since the perturbation values are random values chosen from a uniform distribution, for equal degrees of distortion the increase will not be exactly the same. The error bars represent the standard deviations of the resulting duration and bandwidth of the output signals.

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

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

B. Subjects

Twelve normal-hearing listeners, aged 20–63 years with a mean age of 26 years, participated in the experiment. Pure-tone air-conduction thresholds of the normal-hearing listeners did not exceed 15 dB HL at any octave frequency from 250 to 4000 Hz. In addition, twenty-six sensorineurally hearing-impaired listeners took part in the experiment, aged 24–67 years with a mean age of 48 years.[5] Their intelligibility scores for monosyllabic words in quite were at least 75% correct. The pure-tone, air-conduction threshold in the hearing-impaired listener's better-hearing ear was at least 30 dB HL at one or more frequencies between 250 and 4000 Hz. Thresholds of the better-hearing ear averaged over 0.5, 1, and 2 kHz (the pure-tone average, or PTA) ranged from 17 to 70 dB HL, with a mean PTA of 50 dB HL. All listeners were native Dutch speakers.

C. Stimuli and apparatus

The speech stimuli consisted of sentences and words. The sentence sets contained lists of 13 everyday Dutch sentences of eight to nine syllables read by a female and male speaker (Versfeld *et al.*, 2000). The word sets consisted of lists of balanced meaningful CVC words (Bosman and Smoorenburg, 1995).

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

D. Procedures

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

1. Threshold and UCL

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

Thresholds and UCLs were measured using 1/3-octave noise bands at center frequencies of 250, 500, 1000, 2000, and 4000 Hz. Hearing thresholds were measured using a Békésy tracking (Yantis, 1994) procedure (300-ms noise bursts; repetition rate 2.5 Hz; step size 1 dB). The measurement was ended after 11 level reversals. The average of all but the first reversal level was taken as the hearing threshold. Narrow-band UCLs were measured with 1/3-octave noise bursts that were presented with a 3-dB increase in level for each presentation (300-ms noise burst; repetition rate 1.4 Hz). Listeners were asked to press a button when the noise bursts became uncomfortably loud. Then, the level of the noise burst was immediately diminished by a random amount between 21 and 30 dB, and the ascending procedure was repeated until six responses were obtained. The average of the levels at which the button was pushed was taken as the narrow-band UCL.

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

2. Detection threshold for distortion

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

3. Speech intelligibility

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

After a SRT test using undistorted speech, the SRTa was measured as a function of the degree of distortion (distortion-sensitivity model). The intensity-distortion conditions were 0 (undistorted), 10, and 20 dB. The temporal-distortion conditions were 0 (undistorted), 3, and 7 wavelets. The spectral-distortion conditions were 0, $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{3}{4}$ octave (recall that wavelet phases were distorted in all spectral-distortion conditions).

b. Speech-reception bandwidth threshold (SRBT). In addition to the SRTa, the speech-reception *bandwidth* threshold (SR*B*T) was measured to estimate suprathreshold speech processing. The SR*B*T measure of speech intelligibility was introduced by Noordhoek *et al.* (1999). The SR*B*T is highly sensitive for suprathreshold deficits, as shown in a recent study of Noordhoek *et al.* (2000).

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

E. Speech intelligibility index

To estimate the quality of speech processing of listeners, the SRTa and SR*B*T data were converted to a speech intelligibility index. The speech intelligibility index (SII) (ANSI, 1997) is a physical measure of how much information of speech is available to the listener. The SII correlates highly with speech intelligibility. To perceive speech, normal-hearing listeners need a certain amount of information which can be converted to a SII value. If hearing-impaired listeners need more information, this suggests that their speech processing is degraded. Thus elevated SII values are an indication for a low speech processing quality. The SII model accounts for hearing threshold, self-masking in speech, normal upward spread of masking, and level distortion at high presentation levels. To calculate the SII, speech spectra, noise spectra, and hearing thresholds must be known. As mentioned in Sec. II D 1, hearing thresholds were measured with 1/3-octave noise bands, using Békésy tracking (Yantis, 1994). This procedure probably results in hearing thresholds that are systematically about 4 dB higher than the methods on which the ISO (1961) threshold is based (Noordhoek *et al.*, 2000; Noordhoek *et al.*, 2001). Therefore, in the SII calculations the internal noise level was lowered by 4 dB. The band-importance function for speech material of average redundancy (Pavlovic, 1987) was used.

III. RESULTS AND DISCUSSION

A. Detection thresholds

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

1. Degradation of the intensity accuracy

For the normal-hearing listeners, the detection thresholds for the intensity perturbation, described in Sec. II A 2, ranged from 13 to 23 dB, with a median of 17 dB. For the hearing-impaired listeners, the detection thresholds ranged from 9 to 53 dB, with a median of 18 dB. The overall (normal-hearing plus hearing-impaired listeners: 38 subjects) mean standard error of an individual detection threshold (two measurements) was 3 dB. Individual detection thresholds are shown in Fig. 5(a). This figure will be explained in more detail in Sec. III B. Some hearing-impaired listeners had detection thresholds that were much larger than those for the normal-hearing listeners, but a Mann–Whitney U test showed that the difference in detection threshold between the group of normal-hearing and the group of hearing-impaired listeners was not significant.

2. Degradation of the temporal accuracy

For the normal-hearing listeners, the detection thresholds for temporal perturbation ranged from 0.9 to 1.5 wavelets, with a median of 1.1 wavelets; for the hearing-impaired listeners, the thresholds ranged from 0.6 to 7.4 wavelets, again with a median of 1.1 wavelets. The mean standard error of an individual detection threshold was 0.4 wavelets. Individual detection thresholds are shown in Fig. 5(b). This figure will be explained in more detail in Sec. III B. A few hearing-impaired listeners had detection thresholds that were much larger than those for the normal-hearing listeners, but a Mann–Whitney U test showed that the detection thresholds for the group of hearing-impaired listeners were not significantly higher than those for the group of normal-hearing listeners.

3. Degradation of the spectral accuracy

For the normal-hearing listeners, the detection thresholds for spectral perturbation ranged from 0.22 to 0.39 octave, with a median of 0.26 octave. For the hearing-impaired listeners, the detection thresholds ranged from 0.17 to 1.4 octave, with a median of 0.36 octave. The mean standard error of the individual detection threshold was 0.06 octave. Individual detection thresholds are shown in Fig. 5(c). This figure will be explained in more detail in Sec. III B. A Mann–Whitney U test showed that the detection thresholds for the group of the hearing-impaired listeners were significantly ($p < 0.05$) higher than those for the normal-hearing listeners.

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

B. Suprathreshold speech intelligibility

The aim of this study is to gain insight into the suprathreshold speech processing problems of hearing-impaired listeners. Therefore, speech processing performance was measured by means of the SRTa and SRBT tests. For the normal-hearing listeners, the SRTa ranged from -1.8 to 0.3 dB, with a median of -0.8 dB. For the hearing-impaired listeners, the SRTa ranged from -1.1 to 8.5 dB, with a median of 2.0 dB. The mean standard error of an individual SRTa (six measurements) was 0.7 dB. The hearing-impaired listeners had significantly higher SRTa's than the normal-hearing listeners (Mann–Whitney U test: $p < 0.05$). The SRBT for the normal-hearing listeners ranged from 1.1 to 1.7 octave, with a median of 1.6 octave. The SRBT for the hearing-impaired listeners ranged from 1.5 to 3.4 octave, with a median of 2.1 octave. The standard error of an individual SRBT (two measurements) was 0.3 octave. The hearing-impaired listeners had significantly higher SRBT values than the normal-hearing listeners (Mann–Whitney U test: $p < 0.05$).

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

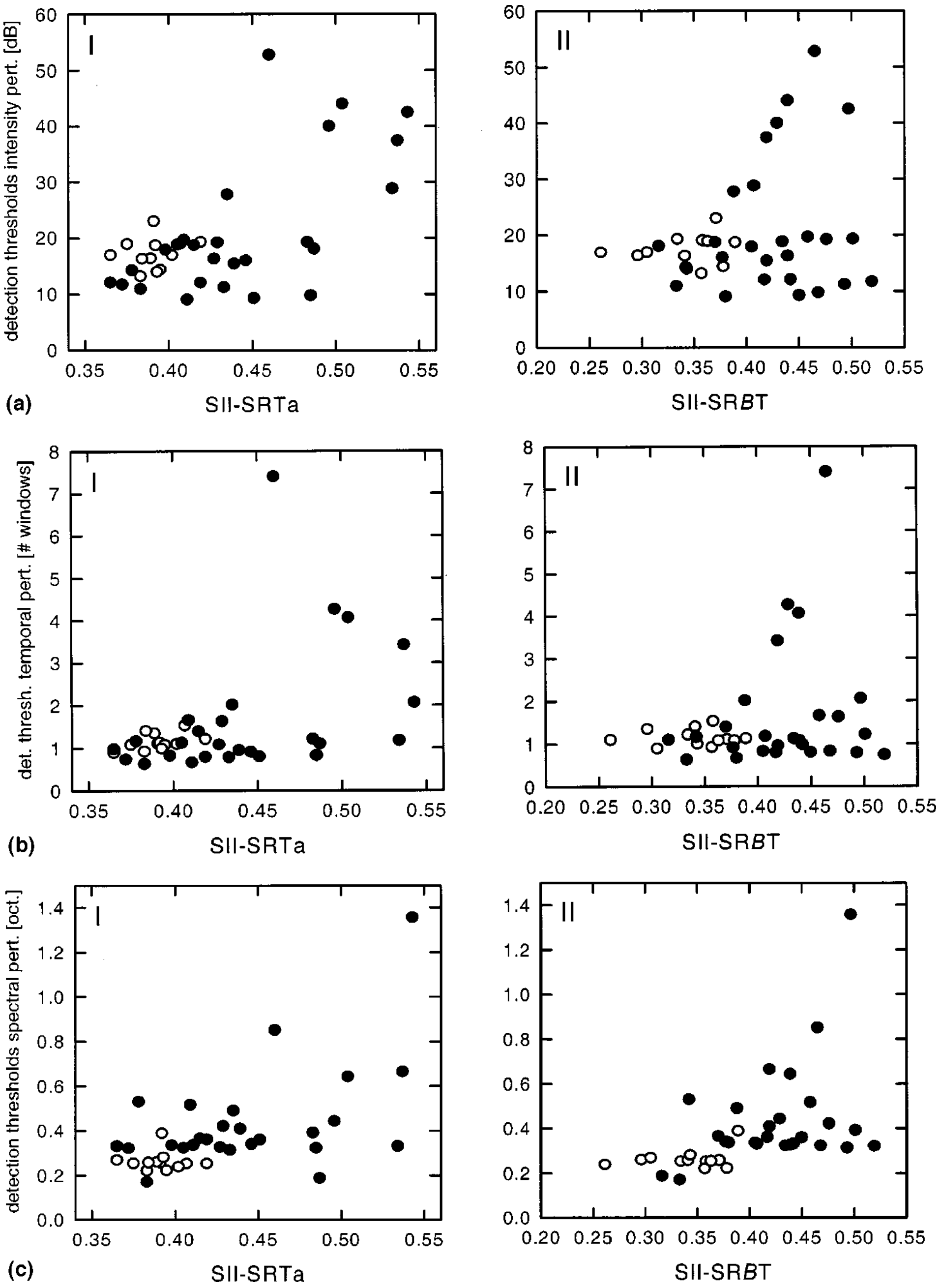

FIG. 5. Individual detection thresholds for intensity (a), temporal (b), and spectral (c) perturbation vs the speech intelligibility index (SII) corresponding with the mean of SRTa scores (panel I) and SRBT scores (panel II) for normal-hearing listeners (open circles) and hearing-impaired listeners (closed circles).

those for the normal-hearing listeners (Mann–Whitney *U* test; $p < 0.05$).

The SII values of the hearing-impaired listeners indicate that their suprathreshold speech processing is clearly distorted. The next step is to explore what aspects of auditory coding are distorted. The detection threshold experiments suggest that hearing-impaired listeners perceive spectral information less clearly than normal-hearing listeners. In Fig. 5(c) the individual detection thresholds for spectral perturbation are plotted as a function of the $\mathrm{SII}_{\mathrm{SRTa}}$ (panel I) and as a function of the $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$ (panel II). Open symbols represent the detection thresholds for the normal-hearing listeners, closed symbols those for the hearing-impaired listeners. The figure shows a correlation between the SIIs and the detection threshold for spectral perturbation. A statistical analysis (Spearman rank correlation) on the combined data for the normal-hearing and hearing-impaired listeners confirmed this: There is a significant ($p < 0.05$) correlation of 0.5 between the detection threshold for spectral perturbation and $\mathrm{SII}_{\mathrm{SRTa}}$, and a significant ($p < 0.05$) correlation of 0.6 between the detection threshold and $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$.

Also correlations between the SIIs and the detection threshold for intensity and temporal perturbation were considered. The individual detection thresholds for intensity and temporal perturbation are plotted in Figs. 5(a) and (b), respectively. Details are the same as in Fig. 5(c). The Spearman rank correlation of the combined data for the normal-hearing and hearing-impaired listeners between the $\mathrm{SII}_{\mathrm{SRTa}}$ and the detection thresholds for intensity and temporal perturbation were significant. Both were 0.4 ($p < 0.05$). The Spearman rank correlations between the $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$ and the detection thresholds for intensity and temporal perturbation were not significant.

Summarizing, a correlation between the detection

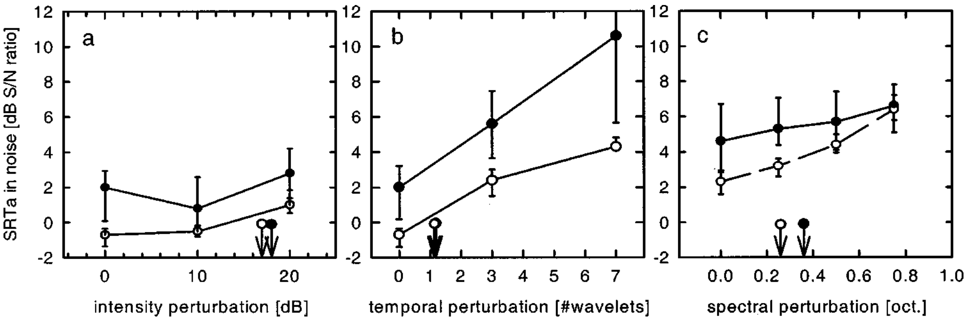

FIG. 6. The median of SRTa values for normal-hearing (open symbols) and hearing-impaired listeners (closed symbols) as a function of distortion. The error bars represent the interquartile ranges. Arrows indicate the median of the detection threshold for each distortion for the normal-hearing listeners (open circle) and hearing-impaired listeners (closed circle): (a) distortion of intensity information; (b) distortion of temporal information; (c) distortion of spectral information.

threshold for the distortion of spectral information and the SII was observed. With respect to the detection thresholds of intensity and temporal perturbation, only the correlation with the $\mathrm{SII}_{\mathrm{SRTa}}$ was significant. Thus distorted processing of spectral information by hearing-impaired listeners relates statistically to their speech-processing deficits. With respect to the processing of intensity and temporal information, this is less clear. In Sec. III C, the relation between the auditory coding accuracy and reduced speech intelligibility is analyzed in a more direct way by means of the distortion-sensitivity model.

C. Distortion-sensitivity model: Group results

Applying the distortion-sensitivity model, the SRTa was measured as a function of the artificial degradation of the spectro-temporal coding of sound, for normal-hearing and hearing-impaired listeners. The results are plotted in Fig. 6. The SRTa is plotted as a function of the degree of distortion of intensity information [panel (a)], temporal information [panel (b)], and spectral information [panel (c)]. Open and closed circles represent the medians of the data for the normal-hearing and hearing-impaired listeners, respectively. The bars represent the interquartile ranges. The arrows represent the medians of the detection thresholds for normal-hearing (open circle) and hearing-impaired listeners (closed circle).

1. Degradation of the intensity accuracy

For all levels of intensity degradation, the hearing-impaired listeners perform poorer than the normal-hearing listeners on the speech intelligibility tests [Fig. 6(a)]. The difference in performance between normal-hearing and hearing-impaired listeners appears to decrease somewhat as a function of the intensity distortion. However, a Mann–Whitney U test showed that this effect was not significant. This is in agreement with the lack of a significant difference in detection thresholds for intensity distortion between normal-hearing and hearing-impaired listeners (Sec. III A 1; medians of the groups represented by arrows). The absence of a difference in sensitivity between normal-hearing and hearing-impaired listeners could be the result from the low perturbation levels used in this study. Higher intensity distortion levels were not measured, because this leads to unwanted spectro-temporal by-products of the signal processing (see van Schijndel *et al.*, 2001). In conclusion, the results do not show a relation between reduced speech intelligibility in noise and a distorted representation of intensity information.

2. Degradation of the temporal accuracy

For all levels of temporal degradation, the medians of the SRTa's for the hearing-impaired listeners are higher than those for the normal-hearing listeners [Fig. 6(b)]. The performances of normal-hearing and hearing-impaired listeners certainly do not converge as a function of temporal perturbation. Instead the performances seem to diverge. This divergence may be related to the fact that the hearing-impaired listeners have less information available from the other, nonperturbed cues. Let's explain this by an example.

Assume as an extreme example that a hearing-impaired listener cannot use the spectral information in speech, but his/her processing of temporal information is as good as that of normal-hearing listeners. When all temporal information is removed from the speech, this hearing-impaired listener cannot understand the speech. The reason is that he/she is deprived of both spectral and temporal cues. Normal-hearing listeners will also be bothered by the removal of temporal information. However, they can still use the spectral information. Thus, looking at the effect of distortion of temporal information on speech intelligibility, performance of this hearing-impaired listener will diverge compared to normal-hearing listeners. However, this divergence does not indicate that this hearing-impaired listener has problems to perceive temporal information. It simply indicates that this hearing-impaired listener has less information available from the other, nonperturbed, cues.

To summarize, the difference in performance between normal-hearing and hearing-impaired listeners does not decrease as a function of temporal perturbation. Actually, the divergence suggests that other cues, not in the temporal domain, are processed less efficiently. In addition, the group of hearing-impaired listeners performed as well as the normal-hearing listeners on the temporal perturbation detection task (Sec. III A 2). In conclusion, the results do not suggest a relation between reduced intelligibility in noise and a distorted representation of temporal information.

3. Degradation of the spectral accuracy

For the most extreme spectral perturbation condition, only the results using the male talker are used, because the male talker was just intelligible in this condition while the female talker was not [see Fig. 6(c)]. The SRTa for the normal-hearing listeners in the spectral reference condition is about 3 dB (median value: 3.1 dB) higher than in the intensity and temporal reference conditions, because the fine structure was perturbed in all spectral conditions (Sec.

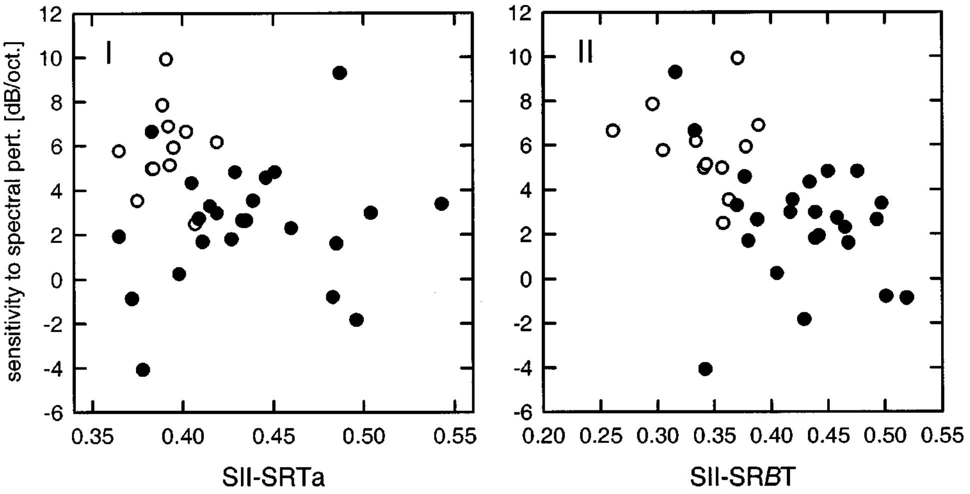

FIG. 7. The individual sensitivities to spectral perturbation for normal-hearing (open symbols) and hearing-impaired listeners (closed symbols) vs the $\mathrm{SII}_{\mathrm{SRTa}}$ (panel I) and vs the $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$ (panel II).

II A 4). The SRTa for the hearing-impaired listeners in the spectral reference condition is about 2.5 dB (median value: 2.4 dB) higher than in the intensity and temporal reference conditions. The difference in SRTa for original speech and speech without fine structure tells something about the role of fine structure. For hearing-impaired listeners this elevation is slightly less than for normal-hearing listeners, but this is not statistically significant. This suggests that perception of fine structure does not relate to speech-perception problems of hearing-impaired listeners.

In the reference condition the median SRTa is higher for the hearing-impaired listeners than for the normal-hearing listeners. When spectral perturbation is applied, the performance for the hearing-impaired listeners converges toward that for the normal-hearing listeners. At $\frac{3}{4}$ octave of spectral perturbation, the performance for the hearing-impaired listeners equals that for the normal-hearing listeners. Mann–Whitney *U* Tests confirm the observed trends: at 0- and $\frac{1}{4}$-octave perturbation the performance for the hearing-impaired listeners is significantly worse than that for the normal-hearing listeners ($p < 0.05$), whereas at $\frac{1}{2}$ and at $\frac{3}{4}$ octave no significant difference exists.

As already shown in Sec. II A, intensity, temporal, and spectral information cannot be manipulated completely independently. Perturbation in one domain will also affect the other domains. However, as shown in Figs. 3 and 4, the spectral perturbation as applied in this study only had a negligible effect on the other domains. Therefore, it seems reasonable to assume that effects present in Fig. 6(c) can be accounted for by a distortion of only the spectral information in speech.

In this study, interdependency of intensity, time, and frequency was only considered from a signal-processing point of view. Also, another interdependency may exist, i.e., within the auditory system. Indeed, a recent study of Loizou *et al.* (1999) demonstrated an interaction between intensity accuracy (expressed in terms of number of quantization levels) and spectral accuracy (expressed in terms of number of spectral channels). This interdependency was not investigated in this study.

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

D. Distortion-sensitivity model: Individual results

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

For intensity information, no correlation between the sensitivity to the distortion and $\mathrm{SII}_{\mathrm{SRTa}}$ or $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$ was observed in the individual data: the Spearman rank correlation between sensitivity to intensity distortion and SII was -0.3 ($p = 0.09$) for the $\mathrm{SII}_{\mathrm{SRTa}}$, and -0.3 ($p = 0.1$) for the $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$.[6] For temporal information, sensitivities were not considered, because of divergence (see Sec. III C 2).

In Fig. 7, the sensitivity to distortion of spectral information is plotted against the individual $\mathrm{SII}_{\mathrm{SRTa}}$ (panel I) and $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$ (panel II). Open symbols represent the data for the normal-hearing listeners; closed symbols those for the hearing-impaired listeners.[7] As is already clear from Fig. 6(c), the median sensitivity of the hearing-impaired listeners is less than that of the normal-hearing listeners. No clear trend between $\mathrm{SII}_{\mathrm{SRTa}}$ and sensitivity is shown [Spearman rank correlation: -0.2 ($p=0.2$)]; however, there is a trend between $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$ and sensitivity: The higher the $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$, the lower the sensitivity to spectral distortion [Spearman rank correlation: -0.6 ($p<0.05$)].[8]

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

In summary, the individual results show a relation between suprathreshold speech processing as quantified by the $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$ and the sensitivity to spectral distortion. This is in agreement with the observed relation between speech processing quality and the detection threshold for spectral perturbation (Sec. III B), and the observed convergence of the performance for normal-hearing and hearing-impaired listeners for increasing degrees of spectral distortion (Sec. III C 3). These results suggest that the auditory processing of spectral information of hearing-impaired listeners is distorted and that this affects speech perception. The poorer the spectral coding, the more problems hearing-impaired listeners have in perceiving speech.

The question remains whether distorted spectral auditory coding is the only cause of suprathreshold speech-processing deficits. A considerable variance is present in the data of Fig. 7. This may be the result of measurement error, but this may also be variance due to factors other than distorted coding of spectral information. By calculating the reliability (Nunnally, 1967) of the variables in the correlation, an estimate of the influence of measurement error can be made. The square root of the product of the reliabilities of two tests gives an estimate of the unsigned maximum correlation possible, given the measurement accuracy.

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

The reliability of the $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$ (two measurements) is 0.7. As a result, the estimate of the unsigned maximum correlation possible between $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$ and sensitivity is 0.5. The correlation observed was -0.6. It may surprise that the absolute value of the observed correlation is larger than the predicted maximum correlation. However, the predicted maximum correlation is only a rough estimate. Therefore, all variance seems explained.

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

E. Comparison to literature

1. Degradation of the intensity accuracy

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

With respect to speech intelligibility as a function of intensity distortion, no significant convergence of the performances for normal-hearing and hearing-impaired listeners was observed. In addition, no significant correlation between the sensitivity to intensity distortion and the SII was found. In contrast, in van Schijndel *et al.* (2001) a significant correlation between sensitivity to intensity distortion and $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$ was observed. Several factors may account for this. Different listener groups were used in the previous and the present study. Since among hearing-impaired listeners a diversity of auditory deficits is observed (see, e.g., Noordhoek *et al.*, 2001), this may lead to a different result. Moreover, although both groups of hearing-impaired listeners had comparable hearing loss, the presentation levels for the second group of listeners was 7 dB lower than for the first group due to lower uncomfortable loudness levels. Due to this difference in dynamic range, the same intensity perturbations may have introduced different loudness perturbations (see van Schijndel *et al.*, 2001). These factors may explain why the correlation in the present study is not significant while in the previous study it was.

2. Degradation of the temporal accuracy

The median detection threshold for temporal distortion of hearing-impaired listeners was not significantly higher than that of normal-hearing listeners. However, some hearing-impaired listeners showed abnormally high detection thresholds. This is in agreement with the literature about temporal resolution. Temporal-resolution deficits occur in some hearing-impaired listeners and not in others (see, e.g., Noordhoek *et al.*, 2001). Whether or not hearing-impaired listeners show temporal-processing deficits also depends on the temporal-resolution test that is used. On some tests of temporal resolution, most hearing-impaired listeners perform as well as normal-hearing listeners (Moore, 1995). Other tests clearly show that hearing-impaired listeners suffer from, for example, excessive forward masking (Festen and Plomp, 1983; Oxenham and Moore, 1995).

The performances of normal-hearing and hearing-impaired listeners did not converge as a function of the distortion of temporal information. This agrees with the study of Duquesnoy and Plomp (1980). They measured how sensitive normal-hearing and hearing-impaired listeners were to reverberation. Reverberation can be considered a very systematic type of distortion of temporal information. The sensitivity of the listeners to reverberation was compared to the Speech Transmission Index (Houtgast and Steeneken, 1973). Their results showed that hearing-impaired listeners were as sensitive to reverberation as normal-hearing listeners.

Based on the previous text, one may conclude that distorted temporal processing is not a factor underlying poor speech intelligibility for the present group of hearing-impaired listeners. This conclusion seems to be in contrast with a recent study of Noordhoek *et al.* (2001). In this extensive study, relations between speech intelligibility and auditory functions in the 1-kHz frequency region were investigated. Results show that a factor related to ''reduced temporal resolution and reduced frequency discrimination seemed to relate to speech-processing deficits.'' As already mentioned (Sec. III D), in the present study not all variance can be explained by distorted spectral processing. Some variance remains unexplained. The underlying factor of this unexplained variance may be distorted temporal (or intensity) processing.

The question remains why in the present study reduced temporal resolution did not show up clearly, while in Noordhoek *et al.*'s study a factor related to temporal resolution and frequency discrimination did. Probably, two differences between Noordhoek's study and the present study may account for this: First, different listener groups were used in Noordhoek's and the present study. As already mentioned in Sec. III E 1, since among hearing-impaired listeners a diversity of auditory deficits is observed, different listener groups may lead to different results. Second, Noordhoek's study concentrated on the 1-kHz frequency region, while the present study looked at the total region from 250 Hz to 4 kHz. Problems related to reduced temporal resolution and/or frequency discrimination may be so frequency specific that looking at a broad frequency range obscures the problem.

3. Degradation of the spectral accuracy

The detection thresholds for spectral distortion were significantly higher for the group of hearing-impaired listeners than for the group of normal-hearing listeners. In addition, convergence of speech-perception performance for normal-hearing and hearing-impaired listeners as a function of spectral distortion was observed. This agrees with the results of Turner *et al.* (1995) that also showed convergence (see Sec. I). It is also in agreement with conclusions of the recent study of Noordhoek *et al.* (2001) that concludes that spectral resolution ''seemed to be related to suprathreshold speech deficits.''

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

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

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

IV. SUMMARY AND CONCLUSIONS

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

The group results showed that the detection thresholds for hearing-impaired listeners with respect to the distortion of intensity and temporal information were not significantly higher than those for normal-hearing listeners. For the distortion of spectral information, the detection thresholds for the hearing-impaired listeners were significantly higher than those for the normal-hearing listeners. Thus hearing-impaired listeners may perceive spectral information less clearly than normal-hearing listeners. With respect to the distortion-sensitivity model, the results (Fig. 6) did not show that the group of hearing-impaired listeners was less sensitive than the group of normal-hearing listeners to intensity and temporal distortion. The group of hearing-impaired listeners was less sensitive than normal-hearing listeners to the distortion of spectral information. Thus the group results suggest that distorted coding of spectral information is an important factor underlying the reduced speech intelligibility observed in hearing-impaired listeners.

Also, the individual results were considered to investigate the relation between reduced speech intelligibility and distorted coding of spectral information in more detail. A significant correlation between the SII, both $\mathrm{SII}_{\mathrm{SRTa}}$ and $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$, and the detection threshold for spectral distortion was observed (Fig. 5). Thus the data reveal a statistical relation between the quality of speech processing, quantified by the SII, and the spectral coding accuracy, quantified by the detection threshold for spectral distortion. In addition, the correlation between the $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$ and the sensitivity to spectral distortion with respect to speech perception was significant (Fig. 7). Thus there is a statistical relation between the quality of speech processing and the effect of distortion of the spectral cues on speech perception. The more pronounced the speech-perception problems of hearing-impaired listeners (in terms of the SII), the less accurate the spectral auditory coding (higher detection thresholds) and the less influence the distortion of spectral information has on speech intelligibility (lower sensitivity to spectral distortion). The individual results support the group result, strongly suggesting that distorted coding of spectral information is the factor underlying the suprathreshold problems encountered by many hearing-impaired listeners when trying to perceive speech.

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

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

(1) The distortion-sensitivity model may be a valuable tool to investigate the underlying causes of reduced speech perception.

(2) Distorted auditory spectral coding may be an important factor underlying the speech-perception problems of hearing-impaired listeners.

(3) Besides distorted coding of spectral information, other factors may play a role in reduced speech intelligibility as well.

ACKNOWLEDGMENTS

This work was supported by The Netherlands Organization for Scientific Research (NWO). The authors wish to thank Ingrid Noordhoek for many stimulating discussions. The authors thank Mary Florentine for useful comments on an earlier version of this paper.

---

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

[2] In van Schijndel *et al.* (2001), the random perturbation factor with which the modulus of each wavelet coefficient was multiplied was chosen from a uniform distribution with boundaries $-L_D$ and $+L_D$.

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

[4] The rms duration of a real function $f(t)$ is defined by $\Delta_t := [1/\|f(t)\|_2]\sqrt{\int_{-\infty}^{\infty}(t-t_0)^2|f(t)|^2 dt}$, with $\|f(t)\|_2 = \sqrt{\int_{-\infty}^{\infty} f^2(t) dt}$ and the center $t_0 := [1/\|f(t)\|_2^2]\int_{-\infty}^{\infty} t|f(t)|^2 dt$. The rms bandwidth is defined analogously [see, e.g., Chui (1992)].

[5] The ages of the group of normal-hearing listeners and the group of hearing-impaired listeners did not match. From the normal-hearing listeners, 10 were in their twenties, 1 was in her thirties, and 1 was in her sixties. From the hearing-impaired listeners, 2 were in their twenties, 9 were in their thirties, 3 were in their forties, 3 were in their fifties, and 9 were in their sixties. In this study, it is assumed that differences in age do not affect the results. This assumption seems reasonable. Literature shows that, for listeners under 70 years of age and with normal hearing, speech intelligibility performance does not vary with age (Studebaker *et al.*, 1997).

[6] Also, an alternative fit was used, the combination of a horizontal line and a sloping line. The subthreshold perturbations (perturbations lower than the detection threshold) were fitted with a horizontal line and all data obtained with larger perturbations with a single sloping line that intercepts the horizontal line at the perturbation threshold. The slope of the sloping line is taken as an ''alternative'' measure of sensitivity. Using these ''alternative'' sensitivities, the correlation with $\mathrm{SII}_{\mathrm{SRTa}}$ is significant [Spearman rank correlation: -0.4 $(p<0.05)$], in contrast with the correlation with the ''original'' sensitivity that was not significant. The correlation between ''alternative'' sensitivity and the $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$ is not significant [Spearman rank correlation: -0.3 $(p=0.07)$].

[7] Negative sensitivities to spectral perturbations may be explained by measurement uncertainty and order/list effects.

[8] The ''alternative'' measure of sensitivity (see footnote 6) leads to the same interpretation of the data. No clear trend between $\mathrm{SII}_{\mathrm{SRTa}}$ and ''alternative'' sensitivity is shown [Spearman rank correlation: -0.1 $(p=0.5)$]; there is a correlation between $\mathrm{SII}_{\mathrm{SR}B\mathrm{T}}$ and sensitivity [Spearman rank correlation: -0.5 $(p<0.05)$].

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

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

ANSI (**1997**). ANSI S3.5-1997, *American National Standard Methods for Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).

Baer, T., and Moore, B. C. J. (**1993**). ''Effects of spectral smearing on the intelligibility of sentences in noise,'' J. Acoust. Soc. Am. **94**, 1229–1240.

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

Buus, S., Florentine, M., and Zwicker, T. (**1995**). ''Psychometric functions for level discrimination in cochlearly impaired and normal listeners with equivalent-threshold masking,'' J. Acoust. Soc. Am. **98**, 853–861.

Chui, K. (**1992**). *Wavelet Analysis and its Applications. Volume 1. An Introduction to Wavelets* (Academic, London), pp. 54–57.

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

Duquesnoy, A. J., and Plomp, R. (**1980**). ''Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis,'' J. Acoust. Soc. Am. **78**, 537–544.

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

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

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

Horst, J. W. (**1987**). ‘‘Frequency discrimination of complex signals, frequency selectivity, and speech perception in hearing-impaired subjects,’’ J. Acoust. Soc. Am. **82**, 874–885.

Houtgast, T. (**1995**). ‘‘Psycho-acoustics and speech recognition of the hearing impaired,’’ in *Proceedings of the European Conference on Audiology* (Noordwijkerhout, The Netherlands), pp. 165–169.

Houtgast, T., and Steeneken, H. J. M. (**1973**). ‘‘The modulation transfer function in room acoustics as a predictor of speech intelligibility,’’ Acustica **28**, 66–73.

ISO (**1961**). International Organization for Standardization, ISO R226-1961, ‘‘Normal Equal-loudness Contours for Pure Tones and Normal Threshold of Hearing under Free Field Listening Conditions’’ (available from American National Standards Institute, New York).

Loizou, P. C., Dorman, M., and Tu, Z. (**1999**). ‘‘On the number of channels needed to understand speech,’’ J. Acoust. Soc. Am. **106**, 2097–2103.

Moore, B. C. J. (**1995**). *Perceptual Consequences of Cochlear Damage* (University Press, Oxford).

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

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

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**1999**). ‘‘Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. I. Normal-hearing listeners,’’ J. Acoust. Soc. Am. **105**, 2895–2902.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**2000**). ‘‘Measuring the threshold for speech-reception by adaptive variation of the signal bandwidth. II. Hearing-impaired listeners,’’ J. Acoust. Soc. Am. **107**, 1685–1696.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**2001**). ‘‘Relations between intelligibility of narrow-band speech and auditory functions, both in the 1-kHz frequency region,’’ J. Acoust. Soc. Am. **109**, 1197–1212.

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

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

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

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

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

Rioul, O., and Vetterli, M. (**1991**). ‘‘Wavelets and signal processing,’’ IEEE Signal Process. Mag. **8**, 14–38.

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

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

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

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gray, G. A. (**1997**). ‘‘Age-related changes in monosyllabic word recognition performance when audibility is held constant,’’ J. Am. Acad. Audiol. Ser. 1 **8**, 150–162.

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

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

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

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

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

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

van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**1999**). ‘‘Intensity discrimination of Gaussian-windowed tones: Indications for the shape of the auditory frequency-time window,’’ J. Acoust. Soc. Am. **105**, 3425–3435.

van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**2001**). ‘‘The effect of intensity perturbations on speech intelligibility for normal-hearing and hearing-impaired listeners,’’ J. Acoust. Soc. Am. **109**, 2202–2210.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). ‘‘Method for the selection of sentence materials for efficient measurement of the speech reception threshold,’’ J. Acoust. Soc. Am. **107**, 1671–1684.

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