A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners

Rhebergen, K.S.; Versfeld, N.J.

DOI
10.1121/1.1861713

Publication date
2005

Published in
The Journal of the Acoustical Society of America

A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners

Koenraad S. Rhebergen[a)]
Department of Clinical and Experimental Audiology, Academic Medical Center, Room D2-223, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Niek J. Versfeld[b)]
Department of Clinical and Experimental Audiology, Academic Medical Center, Room D2-330, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

(Received 1 March 2004; revised 27 December 2004; accepted 31 December 2004)

The SII model in its present form (ANSI S3.5-1997, American National Standards Institute, New York) can accurately describe intelligibility for speech in stationary noise but fails to do so for nonstationary noise maskers. Here, an extension to the SII model is proposed with the aim to predict the speech intelligibility in both stationary and fluctuating noise. The basic principle of the present approach is that both speech and noise signal are partitioned into small time frames. Within each time frame the conventional SII is determined, yielding the speech information available to the listener at that time frame. Next, the SII values of these time frames are averaged, resulting in the SII for that particular condition. Using speech reception threshold (SRT) data from the literature, the extension to the present SII model can give a good account for SRTs in stationary noise, fluctuating speech noise, interrupted noise, and multiple-talker noise. The predictions for sinusoidally intensity modulated (SIM) noise and real speech or speech-like maskers are better than with the original SII model, but are still not accurate. For the latter type of maskers, informational masking may play a role. [DOI: 10.1121/1.1861713]

PACS numbers: 43.71.An, 43.66.Ba, 43.71.Gv, 43.72.Kb [PFA] Pages: 2181–2192

I. INTRODUCTION

In daily life, speech is not always equally intelligible due to the presence of background noise. This noise may mask part of the speech signal such that not all speech information is available to the listener. In order to be able to predict the speech intelligibility under such masking conditions, French and Steinberg (1947), Fletcher and Galt (1950), and later Kryter (1962a, b) initiated a calculation scheme, known as the Articulation Index (AI), which at present still is used by a number of investigators (Rankovic, 1998, 2002; Hogan and Turner, 1998; Müsch and Buus, 2001; Brungart, 2001; Turner and Henry, 2002; Dubno *et al.*, 2002, 2003). In 1984, Pavlovic and others (Dirks *et al.*, 1986; Kamm *et al.*, 1985; Pavlovic, 1984, 1987; Pavlovic and Studebaker, 1984; Pavlovic *et al.*, 1986; Studebaker *et al.*, 1987, 1994) started to re-examine the AI calculation scheme, which has led to a new method accepted as the ANSI S3.5-1997 (1997). Since its revision in 1997, the method is named the Speech Intelligibility Index (SII).

For a given speech-in-noise condition, the SII is calculated from the speech spectrum, the noise spectrum, and the listener's hearing threshold. Both speech and noise signal are filtered into frequency bands. Within each frequency band the factor audibility is derived from the signal-to-noise ratio (SNR) in that band indicating the degree to which the speech is audible. Since not all frequency bands contain an equal amount of speech information (i.e., are not equally important for intelligibility), bands are weighted by the so-called band-importance function. The band-importance function indicates to which degree each frequency band contributes to intelligibility. It depends on the type of speech material involved (e.g., single words or sentences), and other factors. Finally, the SII is determined by accumulation of the audibility across the different frequency bands, weighted by the band-importance function. The resulting SII is a number between zero and unity. The SII can be seen as the proportion of the total speech information available to the listener. An SII of zero indicates that no speech information is available to the listener, an SII of unity indicates that all speech information is available. Model parameters have been chosen such that the SII is highly correlated to intelligibility. The SII model has been developed to predict the *average* speech intelligibility for a given speech-in-noise condition; it does not attempt to predict the intelligibility of the individual utterances (phonemes or words) of a speech fragment. Also, speech redundancy or contextual effects, which are inherent to meaningful speech, are captured in the SII model by choice of the model parameters. Higher speech redundancy simply results in less information (i.e., a lower value for the SII) required for understanding the speech message. Within the context of the present paper, an important observation is that the existing SII model does not take into account any fluctuation in the masking noise, since the SII is computed from the long-term speech and noise spectrum. Therefore, the SII

[a)]Electronic mail: k.s.rhebergen@amc.uva.nl
[b)]Electronic mail: n.j.versfeld@amc.uva.nl

is independent of the amount of fluctuations in the noise signal.

Numerous papers have reported on experiments dealing with speech intelligibility in fluctuating noise. In almost all cases, normal-hearing listeners perform better in conditions with fluctuating noise compared to those with stationary noise of the same rms level (Miller, 1947; Miller and Licklider, 1950; Licklider and Guttman, 1957; de Laat and Plomp, 1983; Duquesnoy, 1983; Festen, 1987, 1993; Festen and Plomp, 1990; Gustafsson and Arlinger, 1994; Bacon *et al.*, 1998; Peters *et al.*, 1998; Brungart, 2001; Versfeld and Dreschler, 2002; Dubno *et al.*, 2002; Nelson *et al.*, 2003). In many cases, this finding has been phenomenologically explained by stating that the listener is ''able to catch glimpses of the speech during the short silent periods of the masking noise'' (Howard-Jones and Rosen, 1992, 1993; Festen, 1993; Peters *et al.*, 1998). Recently, Oxenham and co-workers (Oxenham and Plack, 1997; Plack and Oxenham, 1998; Oxenham *et al.*, 2004) proposed that the nonlinear behavior of the basilar membrane enables increased gain during the silent periods, allowing increased audibility. In hearing-impaired subjects, this nonlinear behavior is less or even absent, which results in decreased audibility during absence of masking noise.

So far, the SII model has been validated only for stationary masking noises, for which it works well. However, it fails to predict speech intelligibility accurately in the case of fluctuating noise maskers (Festen and Plomp, 1990; Houtgast *et al.*, 1992; Versfeld and Dreschler, 2002). Other methods, such as the Speech Transmission Index (STI, Steeneken and Houtgast, 1980), or even the speech-based STI (van Wijngaarden, 2002) also fail at this point. To our knowledge, there is still no method that can predict the speech intelligibility in fluctuating noise accurately. Yet, since most real-life noises do exhibit strong variations over time, there is great interest in a procedure that is able to predict speech intelligibility in fluctuating noises adequately.

In the present paper, an extension to the SII model is proposed in order to be able to predict the speech intelligibility not only in stationary noise, but also in fluctuating noise. The extension consists of an approach where, for a given condition, both speech and noise signal are partitioned into small time frames. Within each time frame, the conventional SII is determined, yielding the speech information available to the listener at that time frame. Next, the SII values of these time frames are averaged, resulting in the SII for that particular noise type. It is hypothesized that this averaged SII is closely related to the speech intelligibility for that condition.

In the next section, an outline of the existing SII model is given. It is followed by a detailed description of the extensions to the existing model, which are introduced to allow predictions of the speech intelligibility in fluctuating noises as well. In extending the SII model, attention has been given to stay as close as possible to the original SII model, thus making as few adaptations as possible. In the choice of the model parameters, this paper concentrates on experiments where speech intelligibility has been assessed with the method of the so-called speech reception threshold (SRT), as described by Plomp and Mimpen (1979). With this method, short everyday sentences are used as speech materials. In Sec. II C the SRT method is described in some detail. Next (in Sec. III) data from the literature are used to evaluate the extended SII model. Finally, in Sec. IV, predictions and limitations of the extended SII model will be discussed.

II. MODEL DESCRIPTION

A. The SII model

A detailed description of the SII model is given in ANSI S3.5-1997 (1997). Here, a brief overview is given so that in the next section the extensions to the existing model are easier to follow.

The SII model basically calculates the average amount of speech information available to a listener. To that extent, the model uses the long-term averaged speech spectrum and the long-term averaged noise spectrum as input. Both speech and noise spectrum are defined as the spectrum level (in dB/Hz) at the eardrum of the listener. Within the model, an option exists to partition the speech and noise spectrum into octave bands, one-third-octave bands, or critical bands. In this paper, spectra are partitioned into critical bands (given in Table I of the ANSI S3.5-1997 standard), although the other two options are equally valid. Within each critical band, the spectrum level is separately determined for both speech and noise. Next, correction factors are taken into account for effects such as upward spread of masking for both speech and noise, inaudibility due to the auditory threshold for pure tones, and distortion due to excessive high speech or noise levels. Then, within each frequency band, the difference between the speech and noise level (signal-to-noise ratio, or SNR) is calculated and this value is multiplied with the so-called band-importance function, which results in the proportion of information in that band that is available to the listener. The band-importance function may depend on the type of speech materials (e.g., sentences or words), or level. Finally, these values are added, yielding the Speech Intelligibility Index (SII), or the amount of speech information available to the listener. For normal-hearing listeners, the SII has proven to be closely related to the average intelligibility in a given condition where speech is masked by a stationary noise masker (Pavlovic, 1987).

B. Extension to the SII model

Since the SII model uses the long-term averaged speech and noise spectrum as input, all temporal characteristics of these signals are lost. As mentioned in the Introduction, large differences in intelligibility exist between masking noises that differ from each other solely with respect to temporal fluctuations (e.g., steady-state versus fluctuating noise). In this section, an extension is presented that does take the temporal characteristics of the masking noise into account. In essence, the SII model is adapted such that the SII is calculated within small time frames, after which the average SII is calculated.

A block diagram of the calculation scheme is presented in Fig. 1. Both speech and noise are analyzed separately for the SII calculation. Although, in principle, regular speech could be used as the speech input signal, speech-shaped

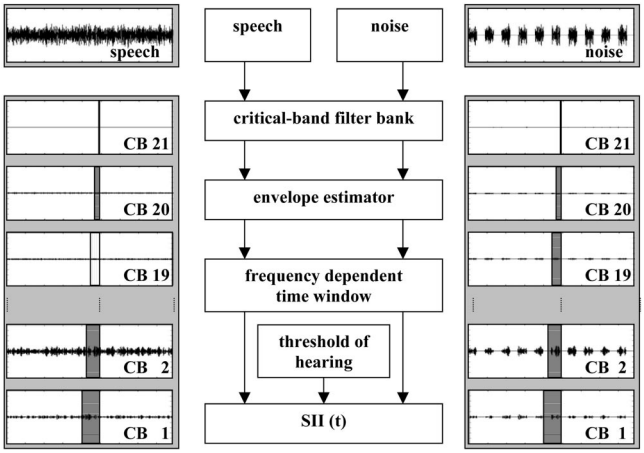

FIG. 1. Schematic overview of the calculation scheme for the extended SII model. A detailed description is given in the main text. The input speech signal (stationary Gaussian noise with the long-term average spectrum of speech) and input noise (in this example interrupted noise with the long-term average spectrum of speech) are separately filtered by a 21 critical-band (CB) filter bank. The envelope of the input speech and noise are estimated in every CB (1–21); the instantaneous intensity is estimated in a frequency-dependent time window, as indicated by the shaded bars (CB1=35 ms to CB21=9.4 ms). Every 9.4 ms an SII is calculated as described by ANSI S3.5-1997. For each of the approximately 200 steps (of 9.4 ms), the instantaneous SII(t) is determined (sentence of about 2 s). Last, the SII for that speech-in-noise condition is determined by averaging across all instantaneous SII(t) values.

noise (i.e., stationary Gaussian noise with the long-term average spectrum of speech) was used. The main reason for this is that, in combination with stationary noise as a noise masker, all SII values are identical to those obtained with the existing SII model. This prerequisite is not easily fulfilled when normal speech signals would be used.

The SII is in principle designed to predict the average intelligibility of speech in noise and not the intelligibility of individual words or phonemes. In any case, the SII is badly defined in case of silent periods occurring within the normal speech signal because, regardless of the masking noise, the SII will always be zero. Thus, even when a speech signal is presented at a clear level without any masking noise, the SII based on regular speech never will reach unity, due to the inherent silent periods in the speech signal. Moreover, problems will occur if one considers the silent periods between sentences. It is clear that large differences in SII may occur when the silent periods between sentences vary, whereas the actual intelligibility should not be different.

The most straightforward approach to determine the SII within small time frames is to window the speech and noise signal at a given point in time, calculate the frequency spectrum (by means of a fast Fourier transform, FFT), and derive an SII from the resulting speech and noise spectrum and the threshold of hearing. However, in order to be able to track the perceptually relevant fluctuations over time, the window length should be small enough. This means that the time window should have a duration of several milliseconds, which is the temporal resolution for normal-hearing listeners based on gap-detection thresholds in the higher frequency bands (Plomp, 1964; Shailer and Moore, 1983, 1987; Glasberg and Moore, 1992; Eddins *et al.*, 1992; Oxenham and Moore, 1994, 1997; Moore *et al.*, 1996; Plack and Oxenham, 1998; Moore, 1997). Unfortunately, such a short time window leads to the signal-analytical problem that the level in the lower frequency bands is not estimated accurately. On the other hand, a longer time window leads to a poorer grasp of the temporal variations of the signal.

It is known that the temporal resolution of the auditory system is frequency dependent (Shailer and Moore, 1983, 1987). Time constants (i.e., integration times) for the lower frequency bands are larger than those for the higher bands. To overcome the analysis problems on the one hand, and to stay close to the characteristics of the auditory system with respect to temporal resolution on the other hand, the signal was first filtered into 21 critical bands, and the window length was chosen to be relatively short in the higher bands and relatively long in the lower bands. Since in the original SII calculations the frequency bands are essentially nonoverlapping (after all, the intensity within each filter band was derived from the frequency spectrum), a FIR filter bank of order 200 [MATLAB function firl(200,Wn)] was used to filter the entire speech and noise signal into the separate bands. Within each band, the temporal envelope was determined by means of a Hilbert transform. At a given time frame, rectangular windows were used with window lengths ranging from 35 ms at the lowest band (150 Hz), to 9.4 ms at the highest band (8000 Hz). These window lengths were taken from Moore (1997, Chap. 4) for gap detection and have been multiplied by 2.5. The factor 2.5 was chosen to provide a good fit to the present data set, as will be discussed below. The windows were aligned such that they ended simultaneously. Within each time frame the intensity was determined, and these, together with the absolute threshold for hearing were used as input to calculate the instantaneous SII, for that given time frame. To calculate the SII, the so-called speech perception in noise (SPIN) weighting function (ANSI S3.5-1997, 1997, Table B.1) was used. This choice seems to be valid, since the speech materials of Plomp and Mimpen (1979) are closely related to the SPIN materials with respect to sentence length and redundancy. Last, the SII for the speech-in-noise condition under consideration was determined by averaging across all instantaneous SII values.

C. Speech reception threshold

In the present paper, the proposed extension to the SII model was evaluated using existing data from the literature. The data differ from each other with respect to a number of variables that all can have an effect on intelligibility, hence on the parameter settings of the SII model. For example, it is known that the type of speech material (monosyllables, words, sentences, etc.), open or closed response set, and native or non-native language acquisition can have a large effect on intelligibility (Bosman and Smoorenburg, 1995; Drullman and Bronkhorst, 2000; van Wijngaarden, 2003). Next, similarity between masker and target, e.g., in the case where both target and masker consist of a male voice (Bronkhorst and Plomp, 1992; Bronkhorst, 2000), has a detrimental effect on the actual threshold (i.e., the signal-to-noise ratio that results in just-intelligible speech). Also, the experimental paradigm influences threshold to a large extent.

The adaptive SRT procedure according to Plomp and Mimpen (1979), and the Just to Follow Conversation (Hygge *et al.*, 1992; Larsby and Arlinger, 1994) result in different threshold levels for the same speech material. Additionally, differences in data acquisition (e.g., strictness of sentence scoring) may have an effect on threshold level. Furthermore, different presentation methods (through headphones, loudspeakers, monaural, binaural, diotic, or dichotic presentation) evidently affect threshold level. If one considers masking noises bearing silent periods, it is likely that, even within a group of normal-hearing subjects, differences in hearing level may affect audibility, and thus intelligibility. Finally, when dealing with spectral differences between masker and target, the method used for calibrating signal levels (e.g., rms, dBA) may have a clear effect.

To enable a comparison between data obtained in different studies, in the present study only thresholds are used that were obtained with the so-called speech reception threshold (SRT) method for sentences, as described by Plomp and Mimpen (1979). Speech materials consist of simple everyday sentences, having a length of 8 to 9 syllables (Plomp and Mimpen, 1979; Nilsson *et al.*, 1994; Versfeld *et al.*, 2000). The SRT is defined as the signal-to-noise ratio (SNR) needed for 50% sentence intelligibility. The SRT is estimated as described by Plomp and Mimpen (1979): A list of 13 sentences, unknown to the listener, is monaurally presented via headphones. The masking noise is presented at a fixed level, whereas the sentence level is varied adaptively. The first sentence starts at a very unfavorable SNR, and is repeated each time at a 4-dB higher level until the listener is able to repeat every word of this sentence exactly. The SNR of the 12 remaining sentences is varied adaptively with a step size of 2 dB using a one-up, one-down procedure. The SNR of the next sentence is increased by 2 dB after an incorrect response and decreased by 2 dB after a correct response. The average adjusted SNR of sentence 4 through 13 is adopted as the SRT for that particular noise condition. With the speech material of Plomp and Mimpen (1979), normal-hearing listeners require an SNR in stationary speech-shaped noise of -5 to -4 dB, which corresponds to an SII between 0.3 and 0.4 (Steeneken, 1992; Bronkhorst, 2000; Noordhoek, 2000; Versfeld and Dreschler, 2002; van Wijngaarden, 2002, 2003). This means that roughly one-third of the speech information is required to the normal-hearing listener (i.e., the SII is between 0.3 and 0.4) to reach the SRT for these sentences.

III. MODEL PREDICTIONS

A. Steady-state speech noise

Speech intelligibility in stationary speech-shaped noise can be well predicted by the existing SII model. There are numerous papers dealing with the SRT in stationary speech noise, and all report for normal-hearing listeners at a fixed noise level between 60 and 80 dBA an SRT for sentences of approximately -4.5 dB (de Laat and Plomp, 1983; Middelweerd *et al.*, 1990; Festen, 1987; Festen and Plomp, 1990; ter Keurs *et al.*, 1993; Versfeld and Dreschler, 2002; Neijenhuis, 2002). For speech in stationary speech noise, an SRT of -4.5 dB results for the existing SII model in an SII value of 0.35.

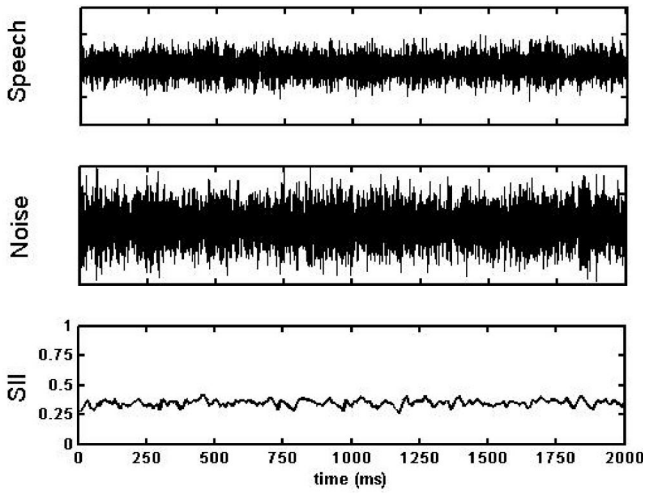

FIG. 2. Representation of the SII with the extended SII model for a speech-in-noise sample of 2 s. The upper panel represents a speech signal of a female speaker. The middle panel represents a stationary speech-shaped masking speech noise. The noise has been scaled to 60 dBA. The target has been scaled to 55.5 dBA, which results in an SNR of -4.5 dB. The lower panel displays the resulting instantaneous SII as a function of time. The SII averaged across time is equal to 0.35.

Figure 2 displays the results of a calculation with the extended SII model for speech in stationary speech noise. The upper panel in Fig. 2 displays the waveform of a speech signal representation (that is—a stationary speech-shaped noise signal instead of an actual speech signal, as discussed in the previous section) with a duration of 2 seconds, presented at a level of 55.5 dBA. Here, speech noise was taken from Versfeld *et al.* (2000) for the female speaker. The middle panel shows a 2-s sample of the stationary speech-shaped noise masker derived from the same female speaker, at a level of 60 dBA. The lower panel in Fig. 2 shows the resulting instantaneous SII, where the SII has been determined every 9.4 ms. Due to the fact that speech and noise signal are uncorrelated (different noise samples), small fluctuations in the instantaneous SII occur. It is easy to see that the SII, averaged across the 2-s sample, is between 0.3 and 0.4. In fact, the average is 0.35, which is identical to the value obtained by the existing SII model. Many conditions with speech in stationary noise have been studied, and all calculations show that neither speech type nor noise type result in differences between the existing SII model and the present extended SII model. In conclusion, the extended SII model yields exactly the same results as the existing SII model, as long as a stationary masking noise is used.

B. Speech noise with a speech-like modulation spectrum

As discussed above, the existing SII model is not able to correctly predict intelligibility for speech in modulated noise. This section deals with speech intelligibility for speech in noise with a speech-like spectrum and a single-speaker modulation spectrum. The generation of this type of noise is described by Festen and Plomp (1990). With normal-hearing

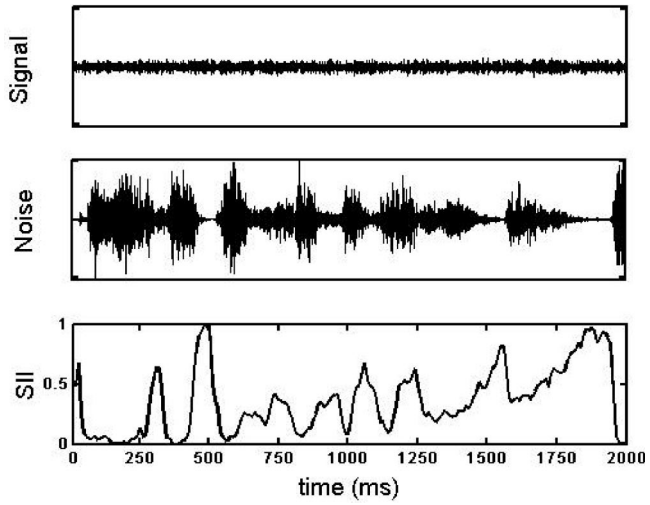

FIG. 3. Representation of the SII with the extended SII model for a speech-in-noise sample of 2 s. The upper panel represents a speech signal of a female speaker. The middle panel represents a fluctuating speech-shaped masking speech noise, as used by Festen and Plomp (1990). The noise has been scaled to 60 dBA. The target has been scaled to 48 dBA, which results in an SNR of −12 dB. The lower panel displays the resulting instantaneous SII as a function of time. The SII averaged across time is equal to 0.35.

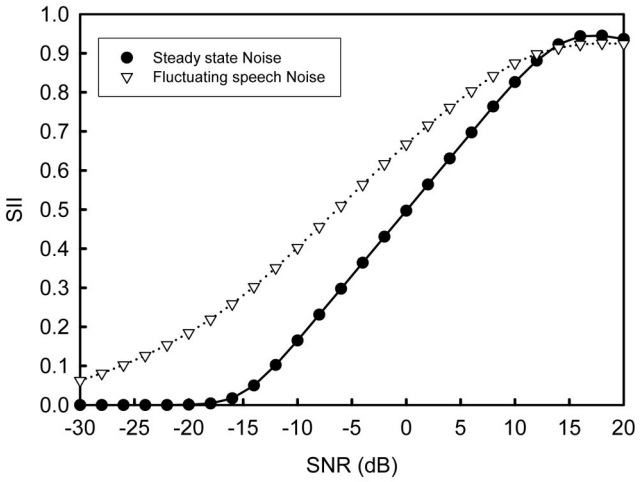

FIG. 4. SII as a function of SNR as calculated with the extended SII model. Filled symbols denote calculations with a stationary noise masker with the long-term spectrum of the female target speaker. Open symbols denote calculations with a fluctuating noise masker with the long-term spectrum of the female target speaker and a speech-like modulation spectrum. The level of the noises was set to 60 dBA.

subjects, several papers report for this condition an SRT around −12 dB (Festen and Plomp, 1990; ter Keurs *et al.*, 1993; Versfeld and Dreschler, 2002; Neijenhuis *et al.*, 2002), when the noise level is between 60 and 80 dBA. Computations with the existing SII model yield a score of 0.089, which is far too low. Figure 3 displays the results of the calculations with the extended SII model, similar to the previous section. The upper panel displays the waveform of a speech signal (again, taken as a stationary speech-shaped noise signal) with a duration of 2 seconds, presented at a level of 48 dBA. The middle panel shows a 2-s sample of the modulated speech noise masker, at a level of 60 dBA. The lower panel in Fig. 3 shows the resulting instantaneous SII, where, in contrast to the findings in Fig. 2, the SII value greatly varies over time. It ranges from values close to zero (at points in time where the speech is entirely masked by the masking noise) to values near unity (at points where the masking noise is momentarily absent). The lower panel thus denotes the amount of speech information available to the listener as a function of time. Averaging across time results in an SII score of 0.35. Because large fluctuations exist over time, a suitably long period has to be chosen to average across. The time interval required to reach stable values for the SII depends on the periodicity, or alternatively, randomness, of the signal as well as on the modulation frequencies in the masking signal. With the present type of masking noise, where the modulations are most prominent near 4 Hz, a period of 2 s appears to be long enough to reach a between-samples standard deviation for the SII of 0.0056. Increasing the period to 4 s decreases the standard deviation of the SII to 0.0030.

Figure 4 displays the SII as a function of the SNR. Here, the masking noise has been kept fixed at 60 dBA, and the level of the speech has been varied between 30 and 80 dBA (thus between SNRs of −30 and +20 dB). With stationary speech noise (denoted as filled symbols in Fig. 4) the SII starts to deviate from zero as the SNR reaches a value of −15 dB and increases almost linearly with the SNR up to a value of +15 dB. At this value, the speech level is about 75 dBA, and the distortion factor in the SII model prevents the SII from reaching unity. The behavior of the SII as a function of SNR with stationary noise is identical for the existing and the extended SII model. Differences between the two models arise when fluctuating noise is used as a masker. Since the existing SII model does not take the amplitude modulations in the noise masker into account, the SII as calculated with the existing SII model will be identical to that calculated for stationary noise. The SII as a function of SNR for fluctuating noise predicted by the extended SII model is given with open symbols in Fig. 4. Even at very low signal-to-noise ratios, there is still some speech information available to the listener and the SII exceeds zero. Increasing the SNR causes the SII to increase, but the slope of the function is not as steep as that calculated for speech in stationary noise. Again, at higher speech levels, the distortion factor of the SII model causes the function to level off, such that the SII does not reach unity. An important observation seen in Fig. 4 is that a constant SII value of 0.35 (the information required to reach threshold) results in an SRT of −4.5 dB for stationary masking noise and −12 dB for fluctuating masking noise.

C. Interrupted speech noise

de Laat and Plomp (1983) measured SRTs for sentences in interrupted (gated) speech noise with a duty cycle of 50%. Modulation frequency was 10 Hz. Masking noise was presented at 65, 75, or 85 dBA. Figure 5 displays the calculations with the extended SII model, similar to Figs. 2 and 3. The upper and middle panel show the speech signal and masking noise signal, respectively. Signal and noise level are 42 and 65 dBA, respectively. The SNR thus is −23 dB. The lower panel shows the SII as a function of time. As seen earlier, the SII is close to zero when the masking noise is present, and is close to unity when the masking noise is absent. Due to the longer integration times in the lower fre-

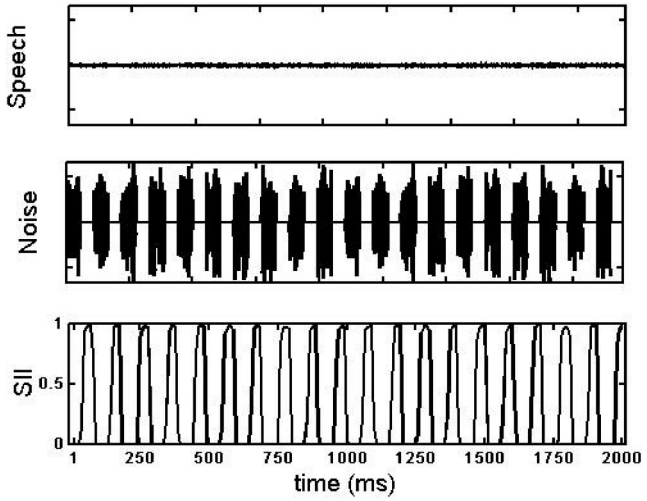

FIG. 5. Representation of the SII with the extended SII model for a speech-in-noise sample of 2 s. The upper panel represents a speech signal of a female speaker. The middle panel represents an interrupted speech-shaped masking speech noise, as used by de Laat and Plomp (1983). The noise has been scaled to 65 dBA. The target has been scaled to 42 dBA, which results in an SNR of −23 dB. The lower panel displays the resulting instantaneous SII as a function of time. The SII averaged across time is equal to 0.35.

quency bands, the SII does not change as rapidly as the interrupted noise, but rather smears out over time. Again, the SII averaged across time is equal to 0.35.

Figure 6 displays the SII as a function of SNR for stationary speech noise (filled symbols), and for the three conditions with 10-Hz interrupted noise used in de Laat and Plomp (1983, open symbols; noise at 65, 75, and 85 dBA). At low SNRs (between −15 and −35 dB), speech is entirely masked at moments when the masking noise is present, and it is audible in the gaps. Due to the gaps in the masking noise, values for the SII are relatively independent of SNR and are still quite large, on the order of 0.3. At even lower SNRs (below −35 dB), SII eventually decreases to zero, due to the fact that the speech signal will fall below the absolute

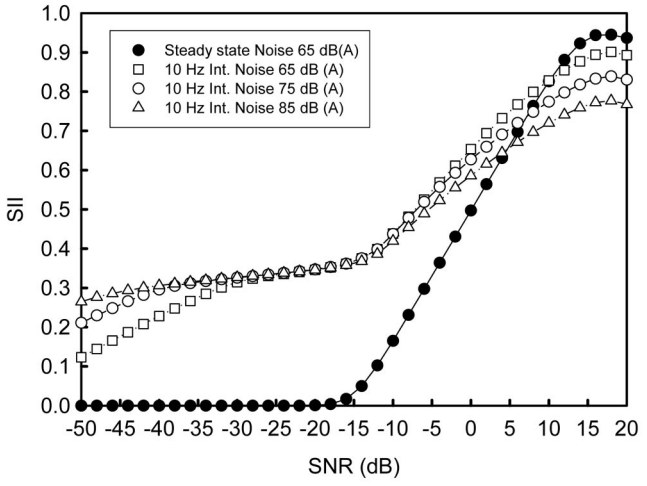

FIG. 6. SII as a function of SNR as calculated with the extended SII model. Filled symbols denote calculations with a stationary noise masker with the long-term spectrum of the female target speaker at a level of 60 dBA. Open squares, circles, and triangles denote calculations with the interrupted noise masker with the long-term spectrum of the female target speaker where the level of the noise was set to 65, 75, and 85 dBA, respectively.

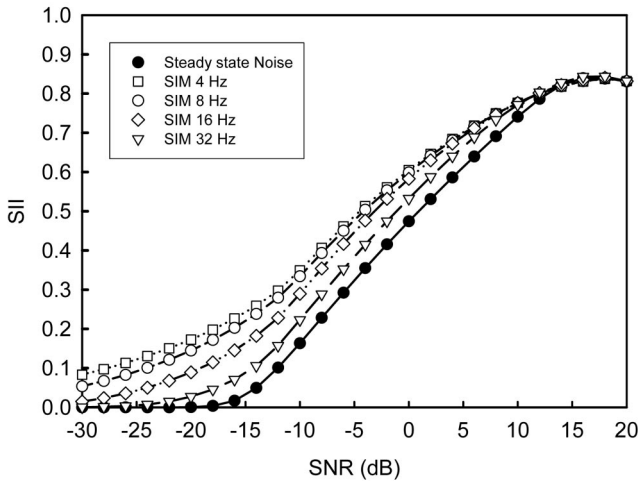

FIG. 7. SII as a function of SNR as calculated with the extended SII model. Filled symbols denote calculations with a stationary noise masker with the long-term spectrum of the female target speaker at a level of 75 dBA. Open squares, circles, diamonds, and triangles denote calculations with SIM noise as a masker at a level of 75 dBA, and a modulation frequency of 4, 8, 16, and 32 Hz, respectively.

threshold. Absolute threshold here has been taken equal to 0 dB (HL). At an SNR of −15 and larger, portions of the speech signal start to exceed the noise signal, and SII increases. Again, at high speech levels, distortion occurs which causes the function to level off. de Laat and Plomp (1983) found an SRT of −23, −26, and −29 dB at a presentation level of the noise of 65, 75, and 85 dBA, respectively. Figure 6 shows that for these conditions a large variation in the SNR causes only a slight variation in the SII. At time frames where the noise signal is present, no speech information is available; but at time frames where the noise masker is absent, the amount of speech information available is determined by the degree of temporal resolution (i.e., forward and backward masking) as well as by the absolute threshold of hearing. Nevertheless, while computations with the existing SII model give an SII of zero, the extended SII model results in values near 0.35.

D. Sinusoidally intensity-modulated speech noise

Festen (1987) measured the SRT for sentences in 100% sinusoidally intensity-modulated (SIM) speech noise. At a presentation level of the noise of 75 dBA he found SRTs of −7.5, −9, −10, −10.2, and −4 dB for modulation frequencies of 4, 8, 16, 32, and ''infinity'' Hz (steady state), respectively. Figure 7 displays the SII as a function of SNR for stationary speech noise (filled symbols), and for four conditions with SIM noise used in the study of Festen (1987, open symbols). Computations with the extended SII model, given an SII of 0.35, result in SRTs of −10, −9, −8, −6.3, and −4 dB for the above-mentioned conditions. The predicted SRT in a 4-Hz SIM noise with the extended SII model seems to be lower compared to SRT values obtained by Festen (1987). Furthermore, the predicted SRT in a 16- or a 32-Hz SIM noise with the extended SII model seems to be higher compared to SRT values obtained by Festen (1987). Although the SRT values obtained with the extended SII model indicate an improvement over the existing model (which pre-

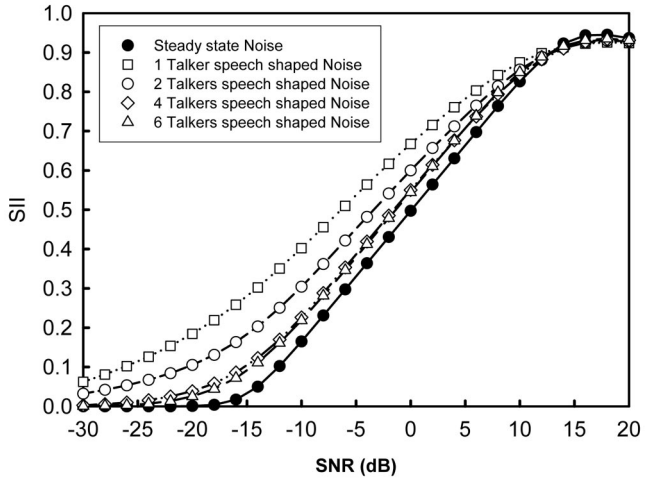

FIG. 8. SII as a function of SNR as calculated with the extended SII model. Filled symbols denote calculations with a stationary noise masker with the long-term spectrum of the female target speaker. Open squares, circles, diamonds, and triangles denote calculations with noise derived from a single, two, four, and six speakers speech-shaped noise. The level of the noises was set to 65 dBA.

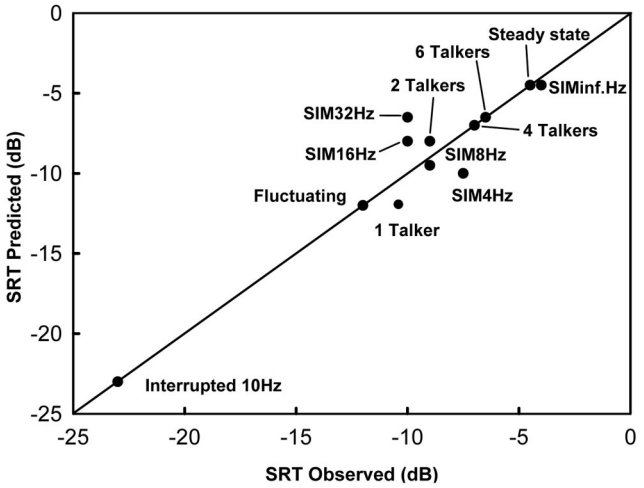

FIG. 9. For a number of different masking noises, the SRT (dB) predicted with the extended SII model is plotted as a function of the observed SRT (dB). Conditions are denoted in short in the figure.

dicts an SRT of −4 dB for all conditions), there are still some deviations. So far, no explanation can be given for this result.

E. Multiple-talker noise

There are numerous papers dealing with the SRT for speech in the presence of one or more competing talkers (e.g., Festen and Plomp, 1990; Bronkhorst and Plomp, 1992; Bronkhorst, 2000; Drullman and Bronkhorst, 2000; Brungart, 2001; Brungart *et al.*, 2001, 2002). It is generally observed that the SRT becomes worse as the number of competing voices increases (Miller, 1947; Carhart *et al.*, 1969; Bronkhorst and Plomp, 1992), eventually resulting in the SRT for stationary speech noise. Bronkhorst and Plomp (1992) measured the SRT for sentences masked by speech-shaped noise modulated by the envelope derived from one, two, four, or six interfering speakers. Observed SRTs were −9.7, −9.9, −7.2, and −6.4 dB, respectively. The stimuli, i.e., speech and fluctuating speech noise, were recorded with a KEMAR manikin and presented monaurally to the subjects. Figure 8 displays for the four conditions of Bronkhorst and Plomp (1992) calculations of the extended SII model as a function of the signal-to-noise ratio where it was attempted to simulate Bronkhorst and Plomps (1992) speech-shaped noises. It shows that at an SII value fixed at 0.35, the SRT increases from −12 dB (for a single interfering speech shaped noise) to −6 dB (for six interfering speech-shaped noises). Although the masking noises were regenerated, since the original masking noises of Bronkhorst and Plomp (1992) were not available, the trend is similar to that reported in the original study.

IV. DISCUSSION

Figure 9 displays the relationship between the observed SRT (i.e., as measured in actual experiments) and the SRT as predicted by the extended SII model for all conditions described in the previous section, as well as some other conditions that will be discussed below. SRTs were calculated by taking the hearing loss fixed at 0 dB(HL) at all audiometric frequencies, and by setting the threshold value of the SII to 0.35. Different SRTs were obtained by taking the associated sample of the masking noise. The diagonal indicates the points where the observed and predicted SRT are equal. Points under the diagonal indicate an overestimation (with respect to performance) of the predicted SRT; points above the diagonal indicate that listeners generally perform better than predicted by the extended SII model. All predicted SRT values are within a few decibels of the diagonal, or even lie on the diagonal, indicating that the model does well with the present set of data. The extended SII model yields a substantial improvement over the existing model. Since the latter is insensitive to modulations in the masking noise, it thus predicts for practically all conditions an SRT of −4.5 dB. The most important finding of this paper is that average speech intelligibility in fluctuating noise can be modeled by averaging the amount of speech information across time.

If the data in Fig. 9 are considered in detail, some of the results obtained with the SIM noises of Festen (1987) seem to deviate to some degree from the diagonal. Festen (1987) found lowest SRTs for modulation frequencies of 16 and 32 Hz. His finding is in contrast with most data from the literature that indicate maximum performance at 10 Hz (Miller and Licklider, 1950; Licklider and Guttman, 1957; Gustafsson and Arlinger, 1994; Trine, 1995; Bronkhorst, 2000; Nelson *et al.*, 2003). The difference in the position of the minimum may be attributable to differences in stimulus type (gated noise versus SIM noise) and speech materials (word versus sentence scoring). There appears to be a large difference in the SRT results (about 16 dB) found by de Laat and Plomp (1983) and Festen (1987) obtained with about the same modulation frequencies [modulation frequency: 10 Hz; SRT: −26 dB for de Laat and Plomp (1983), compared to modulation frequency: 8 Hz: SRT −10 dB for Festen (1987)]. Festen (1987) suggested that this discrepancy can be due to the relatively broad and deep minimum in the inter-

rupted noise compared to that in the SIM noise (Fig. 2 from Festen, 1987). The SRT values, obtained with 16- and 32-Hz SIM noise are very similar, *viz.*, −10 dB, and are 2 to 3 dB better than predicted by the extended SII model. As for now, we have no explanation for this part of Festen's (1987) data. Increasing the modulation frequency of the SIM noise results in gaps that are sufficiently small such that they start to fall within the time window of the extended SII model (i.e., smaller than 35 ms). This results in a decrease in performance, and finally performance will approach that of stationary noise. This condition is indicated by ''SIMinf.Hz'' in Fig. 9, and is close to the diagonal. Decreasing the modulation frequency to 8 Hz also results in a point close to the diagonal. However, a further decrease of the modulation frequency to 4 Hz again results in a deviation from the diagonal. The overestimation of the 4-Hz SIM noise may be accounted for by the fact that with these slow modulation rates, masking of complete words in a sentence can occur. This phenomenon has already been observed by Miller and Licklider (1950), who found optimal performance around modulation rates of 10 Hz. The mere fact that complete words are masked implies that the SRT procedure—where every word of the sentence needs to be repeated correctly—is unsuitable for these low modulation frequencies. Indeed, Trine (1995) shows that in the so-called Just-to-Follow-Conversation (JFC) procedure, the signal-to-noise ratio keeps on decreasing below modulation rates of 8 Hz. In this procedure, the subject is asked to adjust the level of speech in a fixed given noise masker such that he or she is able to ''just follow'' the speech. This procedure does not require the intelligibility of individual syllables, words, or even sentences. Therefore, the optimum performance for 8 Hz is a procedural artifact. Hence, to validate the extended SII model for masking noises comprising modulation rates of, say, 8 Hz and below, procedures other than the SRT procedure of Plomp and Mimpen (1979) should be utilized.

A. Effect of informational masking

The extended SII model may not be able to predict SRTs accurately in conditions where speech and masking noise interfere at a higher level. One example of such interference is when both target speech and masking noise are derived from the same speaker. In that condition, the listener is confused since he or she does not know which signal represents the target and which components of the signal represents the masker. Festen and Plomp (1990) describe a number of conditions where speech is masked by a single speaker or by multiple speakers. Indeed, performance for speech intelligibility in time-reversed masking speech is better than for forward-masking speech. This additional masking, on top of energetic masking, is called informational masking (Bronkhorst, 2000; Brungart, 2001; Brungart *et al.*, 2001): The spoken message of real interfering speech accounts for a rise in SRT.

In another experiment, Festen (1993) measured SRTs in other speech-like maskers. The target speech was uttered by a female speaker (of Plomp and Mimpen, 1979). The interfering speech consisted of comparable sentences from a male voice (Smoorenburg, 1992). In the reference condition, the interfering speech signal consisted of a concatenation of sentences, with no pauses between the sentences. Five other conditions were derived from this reference condition by first dividing the masking speech stream into 2, 3, 6, 12, or 24 separate frequency bands that next were independently shifted in time. One may see this masker as an addition of 2, 3, 4, 6, 12, or 24 speakers where the speech of the individual speakers does not overlap in frequency. The result is a masker that sounds very speech-like. The measured SRTs as well as the SRTs calculated with the extended SII model are displayed in Fig. 10. Different conditions are denoted as B02, B03, B06, B12, and B24, where the number denotes the number of frequency bands. The extended SII model appears to overestimate the observed SRT values of all conditions by 4 to 5 dB. Although speech and noise masker were well discernible, informational masking may have played a role, since the maskers still resembled running speech.

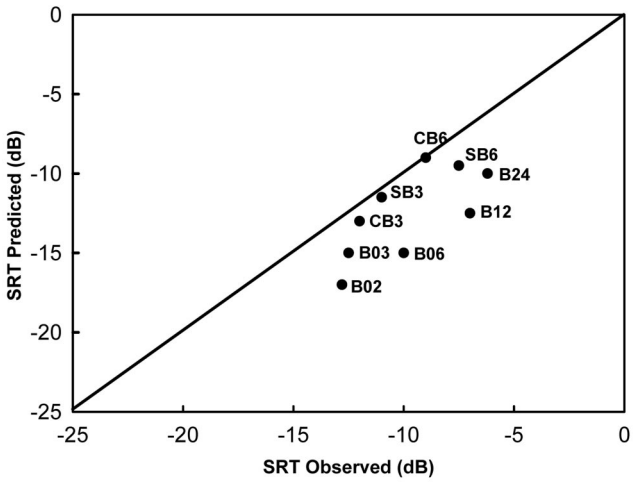

FIG. 10. The SRT (dB) predicted with the extended SII model is plotted as a function of the observed SRT (dB) for the noise maskers used in Festen (1993). Conditions are denoted by abbreviations in the figure. In conditions B02 through B24, conditions consisted of speech fragments that were manipulated by shifting individual frequency bands of the noise masker independently over time. In conditions CB3, CB6, SB3, and SB6, half of the speech masker was replaced by stationary speech noise. For further details the reader is referred to the main text.

In addition to these conditions, Festen (1993) generated other maskers, where the upper 1/3 octave of each frequency band in the 3- and 6-band speech masker was replaced by noise of the same level as the time average of the original masker. Maskers therefore consisted half of stationary speech-shaped noise. The modulated part was either synchronous in time (labeled in Fig. 10 as ''CB'' for ''constant bands'') or shifted in time (labeled in Fig. 10 as ''SB'' for ''shifted bands''). As can be seen in Fig. 10, the extended SII model is able to predict the SRT of all these noise conditions (CB3, CB6, SB3, and SB6) reasonably well, probably due to the fact that the masker is less speech-like.

In summary, when speech-like maskers are used, it is expected that the obtained thresholds are worse than predicted by the extended SII model due to additional (i.e., informational) masking.

B. Steepness of the psychometric function

Festen and Plomp (1990) measured entire psychometric functions for speech in stationary and fluctuating noise. Given the larger dynamic range of fluctuating noise, one would expect a larger range in SNR in which the speech is audible, hence a shallower slope for the fluctuating noise masker. Indeed, with normal-hearing subjects, at the level for which a score of 50% is obtained, Festen and Plomp (1990) found a slope of 21.0%/dB and 11.9%/dB for stationary noise and fluctuating noise, respectively. The present Fig. 4, too, shows a shallower slope for fluctuating noise. With the extended SII model, it is possible to predict the slope of the curve obtained with fluctuating noise from that obtained with stationary noise. To that end, it first should be noted that for SNRs from -9 to -1 dB the psychometric curve with stationary noise in Fig. 6 of Festen and Plomp (1990) ranges from 0% to 100%. Figure 4 shows that this SNR range corresponds to a range for the SII of 0.2 to 0.5. An important observation hence is that within the range of 0.2 to 0.5 of the SII, sentence intelligibility changes from 0% to 100%. Within that range for the SII, both curves in Fig. 4 can be well approximated by a linear function. The curve for stationary noise is given by

$$\mathrm{SII}_S = (15 + \mathrm{SNR}_S)/30, \tag{1}$$

the curve for fluctuating noise is given by

$$\mathrm{SII}_F = (27 + \mathrm{SNR}_F)/40. \tag{2}$$

Festen and Plomp (1990) describe their curves with a logistic function

$$p(\mathrm{SNR}) = \frac{1}{1 + e^{(M - \mathrm{SNR})/S}}, \tag{3}$$

where M is the SNR for which the probability on a correct response $p(\mathrm{SNR})$ is equal to 0.5, and S is the steepness of the function at $p(\mathrm{SNR}) = 0.5$. For the stationary noise curve in Fig. 6 of Festen and Plomp (1990), $M = -4.7$ dB and $S = 1.19$ dB (corresponding to 21.0%/dB as given by Festen and Plomp, 1990). For the fluctuating noise curve, $M = -9.7$ dB and $S = 2.10$ dB (corresponding to 11.9%/dB). The data of Fig. 6 of Festen and Plomp (1990) are replotted in Fig. 11, together with the two functions given by Festen and Plomp (1990), given as solid curves. When $\mathrm{SII}_S = \mathrm{SII}_F$, Eqs. (1) and (2) give the relation between SNR_S and SNR_F

$$\mathrm{SNR}_S = (21 + 3\mathrm{SNR}_F)/4. \tag{4}$$

By insertion of Eq. (4) into Eq. (3), the shape of the function for fluctuating noise is obtained. This curve is plotted as a dotted line in Fig. 11. The predicted curve for fluctuating noise has a slope of 15.6%/dB and a value for M of -13.3 dB. The curve is about 3.8 dB to the left of the data of Festen and Plomp (1990), but has a slope that fits very well to the data of Festen and Plomp (1990), as can be seen when the curve is shifted 3.8 dB to the right, as has been done in Fig. 11 (dashed curve). The slope fits their data even better than their calculated slope of 11.9%/dB. The fact that the calculated curve does not fall on top of the data of Festen and Plomp (1990) is due to the fact that Festen and Plomp (1990) shifted their data to the average results.

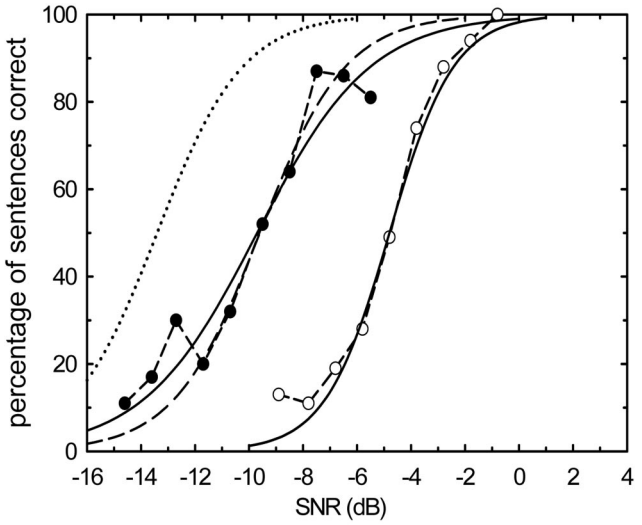

FIG. 11. Percentage of sentences correct as a function of signal-to-noise ratio (dB), for a stationary noise masker (open symbols) and fluctuating noise masker (filled symbols) (replotted from Festen and Plomp, 1990). The two solid curves represent Festen and Plomp's (1990) fit to the data. The dotted curve is predicted by the extended SII model, based on the curve given by Festen and Plomp (1990) for stationary noise. The dashed curve (without symbols) is identical to the dotted curve, except for a shift of 3.8 dB to the right.

C. Effect of absolute threshold

With the calculation of the SII, it was assumed that all subjects had normal hearing; that is, thresholds for all frequencies were taken equal to 0 dB(HL). In real life, thresholds deviate to some degree from this value, but with the normal-hearing group it is generally assumed (ANSI S3.6-1996, 1996) that the hearing level is equal to or less than 15 dB(HL). Given the dynamic range of speech (30 dB) and the presentation level of the masking noise, one can calculate the effect of an elevated threshold. With stationary speech noise as a masker, audibility of average conversational speech starts to play a role only at losses of 50 dB(HL) and larger, as can be calculated with the existing SII model. In contrast, with fluctuating noise and interrupted noise, effects become already noticeable at thresholds of 30 or 15 dB(HL), respectively. The effect of hearing loss on the SII is depicted in Fig. 12 for both a stationary noise masker and an interrupted noise masker. As can be seen in this figure, elevating the threshold from 0 to 15 dB(HL) has no effect on the SII with stationary noise, but has a clear effect with interrupted noise. The two curves with interrupted noise start to overlap near an SNR of -15 dB. For the calculations with the extended SII model, little differences in prediction of the SRT in stationary noise were found by variation of the absolute threshold (HL <50 dB). Figure 12 nevertheless shows that with these fluctuating noise maskers, the effect of absolute threshold can be substantial, especially at lower presentation levels. This could account for the large standard deviation between subjects found by SRT in fluctuating noises (de Laat and Plomp, 1983; Festen, 1987, 1993; Festen and Plomp, 1990; Bronkhorst, 2000; Versfeld and Dreschler, 2002) com-

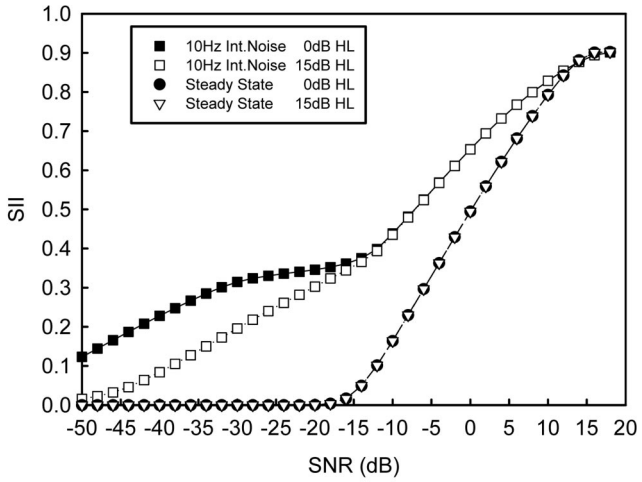

FIG. 12. SII as a function of SNR as calculated with the extended SII model. Filled symbols denote calculations with the absolute threshold set to 0 dB(HL). Open symbols denote calculations with the threshold set to 15 dB(HL). Circles and triangles indicate calculations with a stationary noise masker and squares indicate calculations with interrupted noise masker, respectively, both with the long-term spectrum of the female target speaker. The level of the noises was set to 65 dBA.

pared to the small standard deviation between subjects found by SRT in stationary noises (Plomp and Mimpen, 1979).

D. Effect of window length

With the presentation of the extended SII model, the signals were windowed in time and the length of the time window was frequency dependent. The choice of the time windows was adapted from Moore (1997) and was based on psychophysical data. As discussed above, given these settings, the extended SII model is able to predict the data well. Within a time window, level variations of the signal are averaged. Thus, the longer the time window, the more the signal is smoothed, thus the more the obtained SII will resemble the existing SII (i.e., the SII of stationary noise). On the other hand, if the time windows are taken smaller, all signal variations are caught, which in the case of highly fluctuating maskers as interrupted noise results in better SRTs than actually measured. Calculations have been performed to check whether a single fixed window length for all frequency bands could account for the present data set as well. The results of these calculations show that an optimum fit was obtained with a fixed window length of 12 ms, but that this approach could not account for the data as well as the approach with frequency dependent windows. Yet, it remains possible to manipulate the lengths of the individual time windows, in order to reach an even better fit to the data. However, the present choice of parameters does well, and has the advantage that the window lengths are derived from psychoacoustical measurements. In this paper, rectangular windows have been taken, but future experiments may point to the use of differently shaped windows, such as an exponential window. The latter shape may be more similar to the shape of the forward-masking function.

E. Extensions to the model

In this paper the authors purposely have tried to stay as close as possible to the existing SII model. Extensions to the existing SII model have been proposed, which seem to work well for the SRT with sentences in a given number of noise maskers. To see to what extent the model can be generalized to other types of speech material and noise maskers, measurements should be performed. Although the basic assumptions regarding the extensions may remain valid, it seems plausible that, as with the existing SII model, different speech materials require different weighting functions or window lengths. With the present data set, an SII of 0.35 corresponded to the amount of information required to reach the SRT. These data were obtained with normal-hearing listeners. As discussed extensively by Noordhoek (2000), hearing-impaired subjects often require more speech information to reach threshold, which she attributed to suprathreshold deficits. These deficits probably deal with a decrease in spectral or temporal resolution. With the extended SII model, both decreases in resolution can in principle be modeled by increasing the width of the different frequency bands, or by increasing the window length or window shape. Perhaps more sophisticated adaptations to the SII model [such as the temporal window model of Oxenham (Oxenham and Moore, 1997; Oxenham and Plack, 1997)] are required. It is left to future research to find the extent to which the model is able to describe the data.

F. Other extensions to the SII model

Another shortcoming of the SII model is its inability to account for synergetic and redundant interactions among the various spectral regions of the speech spectrum (Steeneken and Houtgast, 1999; Müsch and Buus, 2001). Due to fact that the SII uses the long-term spectrum of speech and noise (minimum length of 30 s; ANSI S3.5-1997, 1997), these interactions among the various frequency bands are lost. Nevertheless, speech communication is remarkably robust for normal-hearing listeners and does not have to be broadband to be highly intelligible (Allen, 1994; Warren *et al.*, 1995; Lippmann, 1996; Stickney and Assmann, 2000). Steeneken and Houtgast (1999, 2002) implemented a frequency-dependent redundancy correction factor to the STI model, which accounts for synergetic and redundant interactions. Since the STI is related to the SII (van Wijngaarden, 2002), it is in principle possible to implement this redundancy correction factor in the SII calculation method.

V. SUMMARY

The present paper describes an SII-based approach to model SRTs (speech reception thresholds) for sentences masked by fluctuating noise. The basic principle of this approach is that both speech and noise signal are partitioned into small time frames. Within each time frame the instantaneous SII is determined, yielding the speech information available to the listener at that time frame. Next, the SII values of these time frames are averaged, resulting in the SII for that particular noise type. From the literature many SRT values are available for a variety of noise types. In this paper,

it is shown that this approach can give a good account for most existing data. Hence, it forms a valuable extension to the existing SII (ANSI S3.5-1997, 1997) model.

ACKNOWLEDGMENTS

The authors acknowledge Joost Festen and Rob Drullman for providing sound materials of the masking noises used in their papers. Tammo Houtgast, Joost Festen, Gaston Hilkhuysen, and Wouter Dreschler are acknowledged for the stimulating discussions. We are also grateful to Maarten van Beurden, Wouter Dreschler, and Bas Franck for their comments on earlier versions of this paper. We thank the associate editor, Peter Assmann, and the two anonymous reviewers for their detailed constructive comments. Finally, Jan Koopman and especially László Körössy are acknowledged for their help with the computer programming.

Allen, J. B. (**1994**). ''How do humans process and recognize speech,'' IEEE Trans. Speech Audio Process. **2**, 567–577.

ANSI (**1996**). ANSI S3.6-1996, ''American National Standard Methods for Specification for audiometers'' (American National Standards Institute, New York).

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

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

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 impairment,'' Audiology **34**, 260–284.

Bronkhorst, A. W. (**2000**). ''The Cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,'' Acustica **86**, 117–128.

Bronkhorst, A. W., and Plomp, R. (**1992**). ''Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing,'' J. Acoust. Soc. Am. **92**, 3132–3139.

Brungart, D. S. (**2001**). ''Informational and energetic masking effects in the perception of two simultaneous talkers,'' J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., and Simpson, B. D. (**2002**). ''The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal,'' J. Acoust. Soc. Am. **112**, 664–676.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). ''Informational and energetic masking effects in the perception of multiple simultaneous talkers,'' J. Acoust. Soc. Am. **110**, 2527–2538.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). ''Perceptual masking in multiple sound backgrounds,'' J. Acoust. Soc. Am. **45**, 694–703.

de Laat, J. A. P. M., and Plomp, R. (**1983**). ''The reception threshold of interrupted speech for hearing-impaired listeners,'' in *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartman (Springer, Berlin), pp. 359–363.

Dirks, D. D., Bell, T. S., Rossman, R. N., and Kincaid, G. E. (**1986**). ''Articulation index predictions of contextually dependent words,'' J. Acoust. Soc. Am. **80**, 82–92.

Drullman, R., and Bronkhorst, A. W. (**2000**). ''Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation,'' J. Acoust. Soc. Am. **107**, 2224–2235.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2002**). ''Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing,'' J. Acoust. Soc. Am. **111**, 2897–2907.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2003**). ''Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with normal hearing,'' J. Acoust. Soc. Am. **113**, 2084–2094.

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.

Eddins, D. A., Hall, III, J. W., and Grose, J. H. (**1992**). ''The detection of temporal gaps as a function of frequency region and absolute noise bandwidth,'' J. Acoust. Soc. Am. **91**, 1069–1077.

Festen, J. M. (**1987**). ''Speech-perception threshold in a fluctuating background sound and its possible relation to temporal resolution,'' in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Martinus Nijhoff, Dordrecht), pp. 461–466.

Festen, J. M. (**1993**). ''Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice,'' J. Acoust. Soc. Am. **94**, 1295–1300.

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

Fletcher, H., and Galt, R. H. (**1950**). ''The perception of speech and its relation to telephony,'' J. Acoust. Soc. Am. **22**, 89–151.

French, N. R., and Steinberg, J. C. (**1947**). ''Factors governing the intelligibility of speech sounds,'' J. Acoust. Soc. Am. **19**, 90–919.

Glasberg, B. R., and Moore, B. C. (**1992**). ''Effects of envelope fluctuations on gap detection,'' Hear. Res. **64**, 81–92.

Gustafsson, H. A., and Arlinger, S. D. (**1994**). ''Masking of speech by amplitude-modulated noise,'' J. Acoust. Soc. Am. **95**, 518–529.

Hogan, C. A., and Turner, C. W. (**1998**). ''High-frequency audibility: Benefits for hearing-impaired listeners,'' J. Acoust. Soc. Am. **104**, 432–441.

Houtgast, T., Steeneken, H. J., and Bronkhorst, A. W. (**1992**). ''Speech communication in noise with strong variations in the spectral or the temporal domain,'' Proceedings of the 14th International Congress on Acoustics, Vol. 3, pp. H2–6.

Howard-Jones, P. A., and Rosen, S. (**1992**). ''The perception of speech in fluctuating noise,'' Acustica **78**, 258–272.

Howard-Jones, P. A., and Rosen, S. (**1993**). ''Uncomodulated glimpsing in checkerboard noise,'' J. Acoust. Soc. Am. **93**, 2915–2922.

Hygge, S., Ronnberg, J., Larsby, B., and Arlinger, S. (**1992**). ''Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds,'' J. Speech Hear. Res. **35**, 208–215.

Kamm, C. A., Dirks, D. D., and Bell, T. S. (**1985**). ''Speech recognition and the Articulation Index for normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **77**, 281–288.

Kryter, K. D. (**1962a**). ''Methods for the calculation and use of the articulation index,'' J. Acoust. Soc. Am. **34**, 1689–1697.

Kryter, K. D. (**1962b**). ''Validation of the articulation index,'' J. Acoust. Soc. Am. **34**, 1698–1702.

Larsby, B., and Arlinger, S. (**1994**). ''Speech recognition and just-follow-conversation tasks for normal-hearing and hearing-impaired listeners with different maskers,'' Audiology **33**, 165–176.

Licklider, J. C. R., and Guttman, N. (**1957**). ''Masking of speech by line-spectrum interference,'' J. Acoust. Soc. Am. **29**, 287–296.

Lippmann, R. P. (**1996**). ''Accurate consonant perception without mid-frequency speech energy,'' IEEE Trans. Speech Audio Process. **4**, 567–577.

Middelweerd, M. J., Festen, J. M., and Plomp, R. (**1990**). ''Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram,'' Audiology **29**, 1–7.

Miller, G. A. (**1947**). ''The masking of speech,'' Psychol. Bull. **44**, 105–129.

Miller, G. A., and Licklider, J. C. R. (**1950**). ''The intelligibility of interrupted speech,'' J. Acoust. Soc. Am. **22**, 167–173.

Moore, B. C. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, London).

Moore, B. C., Peters, R. W., and Glasberg, B. R. (**1996**). ''Detection of decrements and increments in sinusoids at high overall levels,'' J. Acoust. Soc. Am. **99**, 3669–3677.

Müsch, H., and Buus, S. (**2001**). ''Using statistical decision theory to predict speech intelligibility. II. Measurement and prediction of consonant-discrimination performance,'' J. Acoust. Soc. Am. **109**, 2910–2920.

Neijenhuis, K., Sink, A., Priester, G., van Kordenoordt, S., and van der Broek, P. (**2002**). ''Age effects and normative data on a Dutch test battery for auditory processing disorders,'' Int. J. Audiol. **41**, 334–346.

Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (**2003**). ''Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners,'' J. Acoust. Soc. Am. **113**, 961–968.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). ''Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise,'' J. Acoust. Soc. Am. **95**, 1085–1099.

Noordhoek, I. M. (**2000**). ''Intelligibility of narrow-band speech and its

relation to auditory functions in hearing-impaired listeners,'' Doctoral thesis, Free University, Amsterdam.
Oxenham, A. J., and Moore, B. C. (**1994**). ''Modeling the additivity of nonsimultaneous masking,'' Hear. Res. **80**, 105–118.
Oxenham, A. J., and Moore, B. C. (**1997**). ''Modeling the effects of peripheral nonlinearity in normal and impaired hearing,'' in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Mahwah, NJ), pp. 273–288.
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Oxenham, A. J., Rosengard, P. S., and Braida, L. D. (**2004**). ''Perceptual consequences of normal and abnormal peripheral compression: Potential links between psychoacoustics and speech perception,'' J. Acoust. Soc. Am. **115**, 2421.
Pavlovic, C. V. (**1984**). ''Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment,'' J. Acoust. Soc. Am. **75**, 1253–1258.
Pavlovic, C. V. (**1987**). ''Derivation of primary parameters and procedures for use in speech intelligibility predictions,'' J. Acoust. Soc. Am. **82**, 413–422.
Pavlovic, C. V., and Studebaker, G. A. (**1984**). ''An evaluation of some assumptions underlying the articulation index,'' J. Acoust. Soc. Am. **75**, 1606–1612.
Pavlovic, C. V., Studebaker, G. A., and Sherbecoe, R. L. (**1986**). ''An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals,'' J. Acoust. Soc. Am. **80**, 50–57.
Peters, R. W., Moore, B. C., and Baer, T. (**1998**). ''Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people,'' J. Acoust. Soc. Am. **103**, 577–587.
Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.
Plomp, R. (**1964**). ''Rate of decay of auditory sensation,'' J. Acoust. Soc. Am. **36**, 277–282.
Plomp, R., and Mimpen, A. M. (**1979**). ''Improving the reliability of testing the speech reception threshold for sentences,'' Audiology **18**, 43–52.
Rankovic, C. M. (**1998**). ''Factors governing speech reception benefits of adaptive linear filtering for listeners with sensorineural hearing loss,'' J. Acoust. Soc. Am. **103**, 1043–1057.
Rankovic, C. M. (**2002**). ''Articulation index predictions for hearing-impaired listeners with and without cochlear dead regions,'' J. Acoust. Soc. Am. **111**, 2545–2548.
Shailer, M. J., and Moore, B. C. (**1983**). ''Gap detection as a function of frequency, bandwidth, and level,'' J. Acoust. Soc. Am. **74**, 467–473.
Shailer, M. J., and Moore, B. C. (**1987**). ''Gap detection and the auditory filter: Phase effects using sinusoidal stimuli,'' J. Acoust. Soc. Am. **81**, 1110–1117.
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.
Steeneken, H. J. (**1992**). ''On measuring and predicting speech intelligibility,'' Doctoral thesis, University of Amsterdam.
Steeneken, H. J., and Houtgast, T. (**1980**). ''A physical method for measuring speech-transmission quality,'' J. Acoust. Soc. Am. **67**, 318–326.
Steeneken, H. J., and Houtgast, T. (**1999**). ''Mutual dependence of the octave-band weights in predicting speech intelligibility,'' Speech Commun. **28**, 109–123.
Steeneken, H. J., and Houtgast, T. (**2002**). ''Validation of the revised STI_r method,'' Speech Commun. **38**, 413–425.
Stickney, G. S., and Assmann, P. F. (**2001**). ''Acoustic and linguistic factors in the perception of bandpass-filtered speech,'' J. Acoust. Soc. Am. **109**, 1157–1165.
Studebaker, G. A., Pavlovic, C. V., and Sherbecoe, R. L. (**1987**). ''A frequency importance function for continuous discourse,'' J. Acoust. Soc. Am. **81**, 1130–1138.
Studebaker, G. A., Taylor, R., and Sherbecoe, R. L. (**1994**). ''The effect of noise spectrum on speech recognition performance-intensity functions,'' J. Speech Hear. Res. **37**, 439–448.
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.
Trine, T. D. (**1995**). ''Speech recognition in modulated noise and temporal resolution: Effects of listening bandwidth,'' Unpublished doctoral dissertation, University of Minnesota, Twin Cities, MN.
Turner, C. W., and Henry, B. A. (**2002**). ''Benefits of amplification for speech recognition in background noise,'' J. Acoust. Soc. Am. **112**, 1675–1680.
van Wijngaarden, S. J. (**2002**). ''Past, Present and future of the speech transmission index,'' Proceedings of the International Symposium on STI, TNO Human Factors (Soesterberg, The Netherlands).
van Wijngaarden, S. J. (**2003**). ''The intelligibility of non-native speech,'' Doctoral thesis, Free University, Amsterdam.
Versfeld, N. J., and Dreschler, W. A. (**2002**). ''The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners,'' J. Acoust. Soc. Am. **111**, 401–408.
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.
Warren, R. M., Riener, K. R., Bashford, J. A., and Brubaker, B. S. (**1995**). ''Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits,'' Percept. Psychophys. **57**, 175–182.