Pim van Dijk
Deniz Başkent
Etienne Gaudrain
Emile de Kleine
Anita Wagner
Cris Lanting *Editors*

Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing

Springer Open

Advances in Experimental Medicine and Biology

Volume 894

Advances in Experimental Medicine and Biology presents multidisciplinary and dynamic findings in the broad fields of experimental medicine and biology. The wide variety in topics it presents offers readers multiple perspectives on a variety of disciplines including neuroscience, microbiology, immunology, biochemistry, biomedical engineering and cancer research.

Advances in Experimental Medicine and Biology has been publishing exceptional works in the field for over 30 years and is indexed in Medline, Scopus, EMBASE, BIOSIS, Biological Abstracts, CSA, Biological Sciences and Living Resources (ASFA-1), and Biological Sciences. The series also provides scientists with up to date information on emerging topics and techniques.

<div align="right">2014 Impact Factor: 1.958</div>

More information about this series at http://www.springer.com/series/5584

Pim van Dijk • Deniz Başkent • Etienne Gaudrain
Emile de Kleine • Anita Wagner • Cris Lanting
Editors

Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing

Editors
Pim van Dijk
Deniz Başkent
Etienne Gaudrain
Emile de Kleine
Anita Wagner
Cris Lanting

Department of Otorhinolaryngology / Head
and Neck Surgery
University of Groningen, University
Medical Center Groningen
Groningen
The Netherlands

Speech Intelligibility for Target and Masker with Different Spectra

Thibaud Leclère, David Théry, Mathieu Lavandier and John F. Culling

Abstract The speech intelligibility index (SII) calculation is based on the assumption that the effective range of signal-to-noise ratio (SNR) regarding speech intelligibility is [−15 dB; +15 dB]. In a specific frequency band, speech intelligibility would remain constant by varying the SNRs above +15 dB or below −15 dB. These assumptions were tested in four experiments measuring speech reception thresholds (SRTs) with a speech target and speech-spectrum noise, while attenuating target or noise above or below 1400 Hz, with different levels of attenuation in order to test different SNRs in the two bands. SRT varied linearly with attenuation at low-attenuation levels and an asymptote was reached for high-attenuation levels. However, this asymptote was reached (intelligibility was not influenced by further attenuation) for different attenuation levels across experiments. The −15-dB SII limit was confirmed for high-pass filtered targets, whereas for low-pass filtered targets, intelligibility was further impaired by decreasing the SNR below −15 dB (until −37 dB) in the high-frequency band. For high-pass and low-pass filtered noises, speech intelligibility kept improving when increasing the SNR in the rejected band beyond +15 dB (up to 43 dB). Before reaching the asymptote, a 10-dB increase of SNR obtained by filtering the noise resulted in a larger decrease of SRT than a corresponding 10-dB decrease of SNR obtained by filtering the target (the slopes SRT/attenuation were different depending on which source was filtered). These results question the use of the SNR range and the importance function adopted by the SII when considering sharply filtered signals.

Keywords Speech intelligibility index · Speech in noise · Speech intelligibility measurement

T. Leclère (✉) · D. Théry · M. Lavandier
Laboratoire Génie Civil et Bâtiment, ENTPE, Université de Lyon, Rue Maurice Audin,
69518 Vaulx-en-Velin, France
e-mail: thibaud.leclere@entpe.fr

D. Théry
e-mail: david.thery@entpe.fr

M. Lavandier
e-mail: mathieu.lavandier@entpe.fr

J. F. Culling
School of psychology, Cardiff University, Tower Building, Park Place, Cardiff CF10 AT, UK
e-mail: cullingj@cardiff.ac.uk

1 Introduction

Speech intelligibility in noise is strongly influenced by the relative level of the target compared to that of the noise, referred to as signal-to-noise ratio (SNR). High SNRs lead to better speech recognition than low SNRs. But, although SNRs can take infinite values, it is not the case for speech intelligibility. Some SNR floors and ceilings must exist, such that intelligibility would not be influenced by varying the SNR above or below these limits. Several speech intelligibility models are based on SNRs computations in frequency bands (Rhebergen and Versfeld 2005; Beutelmann et al. 2010; Collin and Lavandier 2013). In the presence of amplitude-modulated or filtered noise, these models could compute infinite SNRs, and then conduct to infinite intelligibility. In order to keep the predictions realistic, a limitation of SNRs must be introduced. Collin and Lavandier (2013) proposed a 10-dB ceiling limit in their model. Other models for intelligibility in modulated noise (Rhebergen and Versfeld 2005; Beutelmann et al. 2010) are based on the speech intelligibility index (SII) calculation (ANSI S3.5 1997) which adopts the SNR values of -15 dB and $+15$ dB as floor and ceiling limits respectively. This range has its origins in the work of Beranek (1947) who reported that the total dynamic range of speech is about 30 dB (in any frequency band) by interpreting the short-term speech spectrum measurements reported by Dunn and White (1940). The aim of this study was to determine floor and ceiling values based on four speech intelligibility experiments and to compare them to those proposed by the SII.

Speech Reception Thresholds (SRTs, SNR yielding 50 % intelligibility) were measured in the presence of a speech target and a speech-shaped noise. Target or noise was attenuated above or below 1400 Hz at different levels of attenuation in order to vary the SNR in the low or high frequency regions. The SRT is expected to increase along with the attenuation level in the case of a filtered target. Conversely, when the noise is filtered, the SRT should decrease while the attenuation level is increased. In both cases, SRTs are expected to remain unchanged and form an asymptote beyond a certain attenuation level: variations of SNR should not influence speech intelligibility any longer. The floor and ceiling values (attenuation from which SRTs are no longer influenced) obtained in each experiment will be compared to those adopted by the SII standard $[-15$ dB$; +15$ dB$]$.

General methods of the experiments are presented first, detailing the conditions and stimuli tested in this study, then followed by the results of each experiment. These results are finally discussed in the last section.

2 Methods

2.1 Stimuli

2.1.1 Target Sentences

The speech material used for the target sentences was designed by Raake and Katz (2006) and consisted of 24 lists of 12 anechoic recordings of the same male voice

digitized and down-sampled here at 44.1 kHz with 16-bit quantization. These recordings were semantically unpredictable sentences in French and contained four key words (nouns, adjectives, and verbs). For instance, one sentence was "la LOI BRILLE par la CHANCE CREUSE" (the LAW SHINES by the HOLLOW CHANCE).

2.1.2   Maskers

Maskers were 3.8-s excerpts (to make sure that all maskers were longer than the longest target sentence) of a long stationary speech-shaped noise obtained by concatenating several lists of sentences, taking the Fourier transform of the resulting signal, randomizing its phase, and finally taking its inverse Fourier transform.

2.1.3   Filters

Digital finite impulse response filters of 512 coefficients were designed using the host-windowing technique (Abed and Cain 1984). High-pass (HP) and low-pass (LP) filters were used on the target or the masker at different attenuations (0 to 65 dB) depending on the experiment. The cut-off frequency was set to 1400 Hz for both HP and LP filters to achieve equal contribution from the pass and stop bands according to the SII band importance function (ANSI 1997; Dubno et al. 2005).

2.2   Procedure

Four experiments were conducted to test each filter type (HP or LP) on each source (target or masker). Except for experiment 1 (HP target), each experiment was composed of two sub-experiments of eight conditions because no asymptote was reached with the first set of eight conditions. Experiment 1 tested only eight conditions. In each sub-experiment, each SRT was measured using a list of twelve target sentences and an adaptive method (Brand and Kollmeier 2002). The twelve sentences were presented one after another against a different noise excerpt corresponding to the same condition. Listeners were instructed to type the words they heard on a computer keyboard after each presentation. The correct transcript was then displayed on a monitor with the key words highlighted in capital letters. Listeners identified and self-reported their score (number of correct key words they perceived). For the first sentence of the list, listeners had the possibility to replay the stimuli, producing an increase in the broadband SNR of 3 dB, which was initially very low (-25 dB). Listeners were asked to attempt a transcript as soon as they believed that they could hear half of the words in the sentence. No replay was possible for the following sentences, for which the broadband SNR was varied across sentences by modifying the target level while the masker level was kept constant at 74 dB$_A$ SPL. For a given sentence, the broadband SNR was increased if the score obtained at the previous sentence was greater than 2, it was decreased if the score was less than 2 and it remained unchanged if the score was 2. The sound level of

the kth ($2 < k < 12$) sentence of the list (L_k, expressed in dB SPL) was determined by Eq. 1 (Brand and Kollmeier 2002):

$$L_k = L_{k-1} - 10 \times 1.41^{-i} \times ((SCORE_{k-1} / 4) - 0.5)$$

where $SCORE_{k-1}$ is the number of correct key words between 1 and 4 for the sentence k-1 and i is the number of times $(SCORE_{k-1}/4)$-0.5 changed sign since the beginning of the sentence list. The SRT was taken as the mean SNR in the pass band across the last eight sentences. In each sub-experiment, the SRT was measured for eight conditions presented in a pseudorandom order, which was rotated for successive listeners to counterbalance the effects of condition order and sentence lists, which were presented in a fixed sequence. Each target sentence was thus presented only once to every listener in the same order and, across a group of eight listeners, a complete rotation of conditions was achieved. In each experiment, listeners began the session with two practice runs, to get used to the task, followed by eight runs with break after four runs.

2.3 Equipment

Signals were presented to listeners over Sennheiser HD 650 headphones in a double walled soundproof booth after having been digitally mixed, D/A converted, and amplified using a Lynx TWO sound card. A graphical interface was displayed on a computer screen outside the booth window. A keyboard and a computer mouse were inside the booth to interact with the interface and gather the transcripts.

3 Listeners

Listeners self-reported normal hearing and French as their native language and were paid for their participation. Eight listeners took part in each sub-experiment. Within each experiment, no listener participated at both sub-experiments since the target sentences used in each sub-experiment were the same.

4 Results

4.1 HP Target

Figure 1 presents the SRTs measured in the presence of a HP-filtered target as a function of the filter attenuation in the low-frequency region. SRTs first increased linearly from 0 to 15-dB attenuation and then remained constant for further attenu-
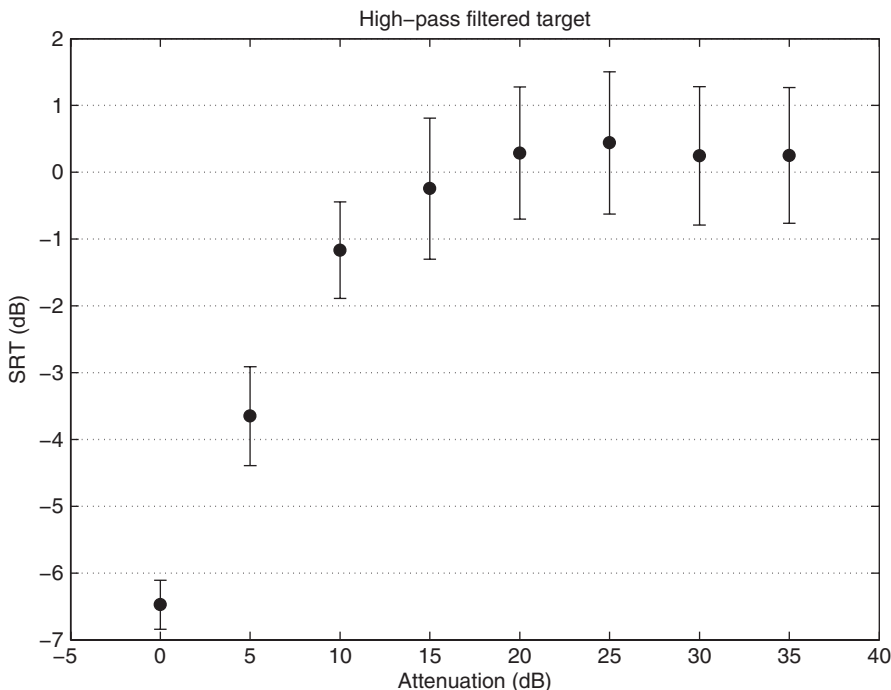
Fig. 1 SRT measurements for a high-pass filtered target as a function of the filter attenuation

ations. Intelligibility was not disrupted any longer after filtering out the target by 15-dB attenuation. A one-factor repeated measures analysis of variance (ANOVA) was performed on the experimental data, showing a main effect of the filter attenuation [$F(7,7) = 10.58$; $p < 0.001$]. Tukey pairwise comparisons were performed on the data: none of the SRTs from 10-dB attenuation to 35-dB were significantly different from each other. By fitting a broken stick function on the experimental data, the floor value was determined at 13-dB attenuation and the slope of the linear increase of SRT was 0.53 dB SRT/dB attenuation. The same fitting process has been used in each experiment to determine the slope and the floor (or ceiling) value.

4.2   LP Target

SRT measurements in the presence of a LP-filtered target are plotted in Fig. 2 as a function of the filter attenuation. As in the HP case, SRTs increased linearly with a slope of 0.48 dB SRT/dB attenuation, but unlike the HP case, SRTs kept increasing until a floor value of 37-dB attenuation. Intelligibility remained at a SRT of about 10 dB for further attenuations. A one-factor repeated measures ANOVA was performed on each sub-experiment independently. In both sub-experiments, a main
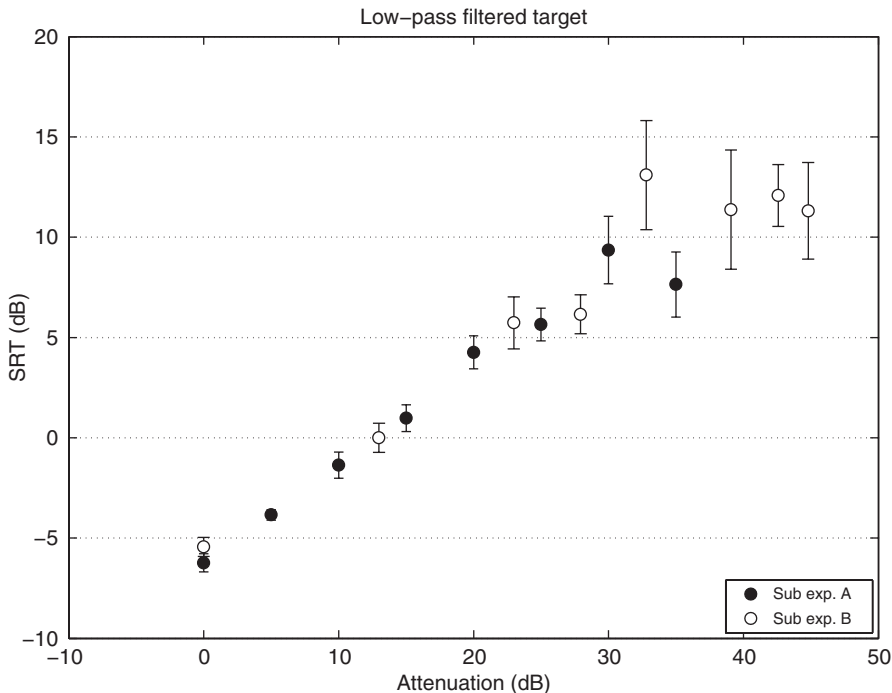
Fig. 2 SRT measurements for a low-pass filtered target as a function of the filter attenuation. Filled circles correspond to the first sub-experiment while open circles correspond to the second

effect of attenuation was observed [$F(7,7) > 16.2$; $p < 0.01$]. Post-hoc Tukey pairwise comparisons were performed on the data of each sub-experiment. In sub-experiment A (filled circles), the four SRTs at the highest level of attenuation (20, 25, 30 and 35 dB) are not significantly different from each other. In sub-experiment B (open circles), the six SRTs at the highest attenuation level (23, 28, 33, 39, 42 and 45 dB) are not significantly different from each other.

4.3 HP Masker

SRTs measured with a HP filtered masker are presented as a function of the filter attenuation in Fig. 3. SRTs decreased linearly with a slope of −0.65 dB SRT/dB attenuation indicating an improvement of speech intelligibility by filtering out the low frequencies in the masker signal. At 43-dB attenuation (ceiling value), SRTs stopped decreasing and presented an asymptote at about −35 dB. A one-factor repeated measures ANOVA indicated a significant main effect of the filter attenuation on speech intelligibility [$F(7,7) > 41.8$; $p < 0.01$ in each sub-experiment]. Tukey pairwise comparisons were performed on the dataset of sub-experiments A and B.
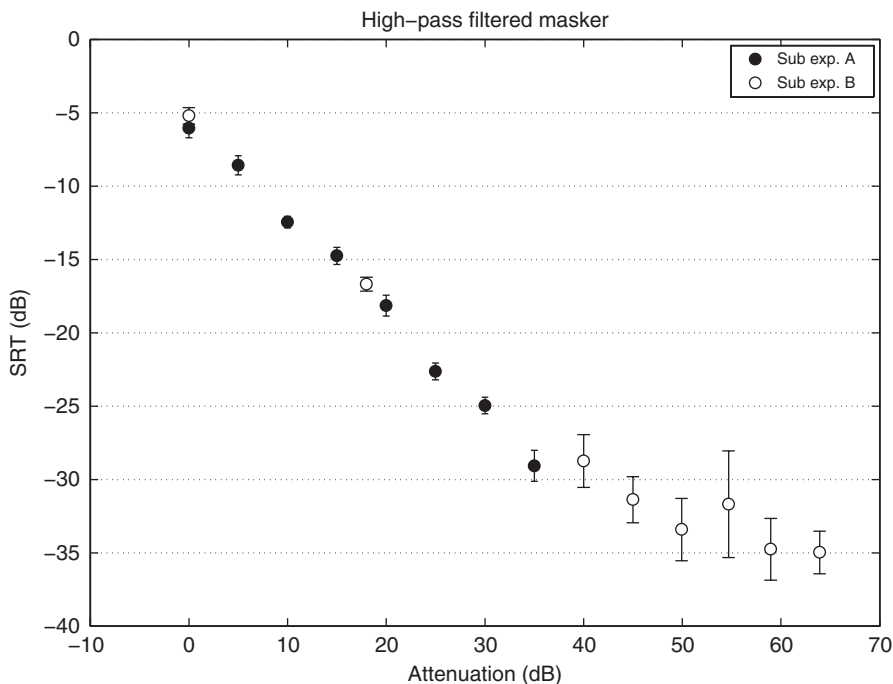
Fig. 3 SRT measurements for high-pass filtered masker as a function of the filter attenuation. Filled circles correspond to the first sub-experiment while open circles correspond to the second

In sub-experiment A (filled circles), only the pairs of SRTs at the attenuations 30/25, 10/15 and 0/5 were not significantly different. All the other pairs of SRTs were significantly different from each other. In sub-experiment B (open circles), SRTs for 0 and 20-dB attenuation were significantly different from each other and all SRTs obtained for 40-dB attenuation at 40 dB and above were not significantly different from each other.

4.4 LP Masker

Figure 4 presents the SRTs measurements obtained in the presence of a LP filtered masker as a function of the filter attenuation. As in the HP case, SRTs linearly decreased with attenuation until the ceiling value of 36-dB attenuation. For further attenuations, SRTs were constant at about −35 dB. The slope of the linear decrease of SRTs was −0.76 dB SRT/dB attenuation. A one-factor repeated measures ANOVA was performed on the data from each sub-experiment independently. A main effect of the filter attenuation was found [$F(7,7) > 44.5$; $p < 0.01$], which was further investigated by performing Tukey pairwise comparisons on the dataset from each sub-experiment. In sub-experiment A (filled circles), all pairs of SRTs were differ-
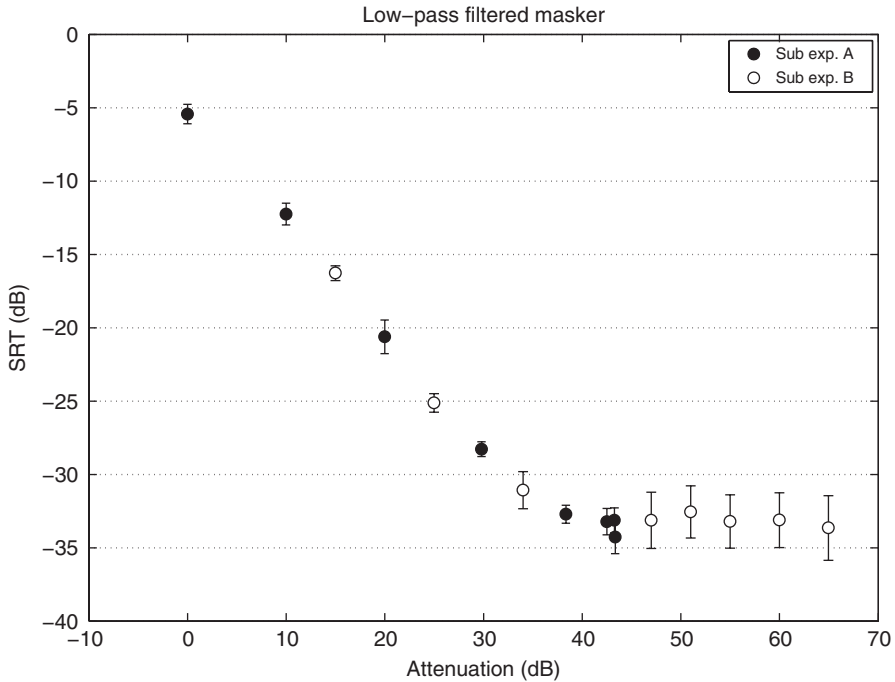
Fig. 4 SRT measurements of low-pass filtered masker as a function of the filter attenuation. Filled circles correspond to the first sub-experiment while open circles correspond to the second

ent from each other except for those corresponding to an attenuation at 38 dB and above. In sub-experiment B (open circles), none of the SRTs between 34-dB and 65-dB attenuation were significantly different. SRTs obtained for lower attenuations were all significantly different from each other.

5 Discussion

In each experiment, speech intelligibility was strongly influenced by the SNR in a specific band. However, SRTs were not affected any longer beyond a certain attenuation level (floor or ceiling values) and remained constant. The floor/ceiling values observed in these experiments are not in agreement with the SII assumptions, except when the target was HP-filtered. In this specific case, the results suggested a floor value at 13-dB attenuation. The results from the other experiments suggested a larger range of SNRs contributing to speech intelligibility $[-37 \text{ dB}; 43 \text{ dB}]$. When the target is filtered, the floor value seemed to be frequency-dependent in contrast to the SII approach which proposes the same SNR range in each frequency band. It is worth noting that the SII standard was not designed for sharply filtered speech or sharply filtered noise such as the stimuli used in the present study. The SII is then

not expected to predict the data presented here but other values for the ceiling and floor SNRs than those proposed by the standard are needed to describe speech intelligibility of sharply filtered signals.

Changing the filter type (HP or LP) had a small influence on the slope of the linear increase (or decrease) of SRTs before reaching the asymptote, confirming that the chosen low and high frequency regions contributed equally to speech intelligibility. But, the results showed that filtering the masker seemed to have a larger impact on speech intelligibility than filtering the target. SRTs were not symmetrical as the SNR was varied up or down with the same amount. The benefit was greater when the filtered part of the noise decreased rather than when the filtered part of the target increased. This result questions the uniformly distributed importance over the $[-15; +15]$ interval adopted by the SII and suggests that a greater importance should be attributed to positive SNRs compared to negative SNRs.

Studebaker and Sherbecoe (2002) derived intensity importance functions from intelligibility scores in five frequency bands. The functions they obtained differed from the one used in the SII calculation. Their functions are frequency-dependent and nonlinear along SNRs, which is in agreement with the findings of the present study. However, the derivation of their results yielded importance functions defined between SNRs of -15 dB and $+45$ dB. The results observed here suggested a floor SNR of -13 dB only when the target was HP-filtered. For the other conditions, lower SNRs need to be considered. The importance function of Studebaker and Sherbecoe (2002) also attributes negative contribution to very high SNRs (>30 dB on average across frequency bands) regarding speech intelligibility. This negative contribution might be due to high sound levels used in their study which could have led to poorer word recognition scores (Dubno et al. 2005). Their derived importance functions take into account both the effect of SNR and absolute speech level. It would be preferable to separate the influence of these two parameters. Their approach could however be inspiring in order to propose simple frequency-dependent SNR-importance functions, allowing modelling the SRTs measured in the present study.

6   Conclusion

Speech reception thresholds were measured in four speech intelligibility experiments. Target or noise was either high-pass filtered or low-pass filtered with different SNRs in the rejected band by varying the attenuation level of the filter. As expected, it was observed in each experiment that SRTs remained constant beyond a certain attenuation level (i.e. a certain SNR in the rejected band). In general, the SNR value from which SRT was not influenced any longer differed from previous values reported in the literature, especially in the SII standard. These results provide ceiling and floor values of SNR for wide frequency bands based on experimental measurements. They do not question the validity of the SII (which was not designed for sharply filtered sources) but they rather point out the need of nonlinear SNR-

importance functions in speech intelligibility models based on SNR weightings to predict SRTs, especially if they aim to account for filtered sources. Further work needs to be done to determine ceiling and floor values in narrower frequency bands. To model these results, SNR-importance functions also need to be built, for example following the approach of Studebaker and Sherbecoe (2002) who proposed a two-dimensional importance function.

References

Abed A-EHM, Cain GD (1984) The host windowing technique for FIR digital filter design. IEEE Trans Acoust Speech Signal Process (ASSP) 32:683–694

Beranek LL (1947) The design of speech communication systems. Proc Inst Radio Eng 35:880–890

Beutelmann et al (2010) Revision, extension, and evaluation of a binaural speech intelligibility model. J Acoust Soc Am 127:2479–2497

Brand T, Kollmeier B (2002) Efficient adaptive procedures for thresholds and concurrent slope estimates for psychophysics and speech intelligibility tests. J Acoust Soc Am 111(6):2801–2810

Collin B, Lavandier M (2013) Binaural speech intelligibility in rooms with variations in spatial location of sources and modulation depth of noise interferers. J Acoust Soc Am 134:1146–1159

Dubno JR, Horwitz AR, Ahlstrom JB (2005) Recognition of filtered words in noise at higher-than-normal levels: decreases in scores with and without increases in masking. J Acous Soc Am 118(2):923–933

Dunn HK, White SD (1940) Statistical measurements on conversational speech. J Acoust Soc Am 11:278–288

Raake A, Katz BFG (2006). SUS-based method for speech reception threshold measurement in French. Proc Lang Resour Eval Conf 2028–2033

Studebaker GA, Sherbecoe RL (2002) Intensity-importance functions for bandlimited monosyllabic words. J Acoust Soc Am 111:1422–1436

Versfeld R (2005) A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. J Acoust Soc Am 117:2181–2192

ANSI S3.5 (1997). Methods for calculation of the speech intelligibility index. American National Standards Institute, New York