NEAR END LISTENING ENHANCEMENT OPTIMIZED WITH RESPECT TO SPEECH INTELLIGIBILITY INDEX

Bastian Sauert and Peter Vary

Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, 52056 Aachen, Germany
email: {sauert,vary}@ind.rwth-aachen.de, web: www.ind.rwth-aachen.de

ABSTRACT

Signal processing algorithms for near end listening enhancement allow to improve the intelligibility of clean (far end) speech for the near end listener who perceives not only the far end speech but also ambient background noise. A typical scenario is mobile communication conducted in the presence of acoustical background noise such as traffic or babble noise.

In this contribution we analyze the calculation rules of the Speech Intelligibility Index (SII) and derive a simple condition for the speech spectrum level of every subband that maximizes the SII for a given noise spectrum level. This rule is used to derive a theoretical bound for a maximum achievable SII as well as a new SII optimized algorithm for near end listening enhancement. The impact of ignoring masking effects in the algorithm is also investigated and seconds our SNR recovery algorithm proposed earlier.

Instrumental evaluation shows that the new algorithm performs close to the established theoretical bound.

1. INTRODUCTION

Mobile communication is often conducted in the presence of acoustical background noise such as traffic or babble noise. This leads to the problem that the *near end* listener perceives a mixture of the clean *far end* speech and the acoustical background noise from the *near end* and thus experiences a reduced speech intelligibility.

For the problem of near end listening enhancement, in contrast to the problem of noise reduction, the noise signal cannot be influenced because the person is located in a noisy environment and the noise reaches the ear with hardly any possibility to intercept. Therefore, a reasonable approach to improve intelligibility by digital signal processing is to manipulate the *far end* speech signal in dependence of the *near end* background noise.

In [1], we proposed an approach which amplifies the far end speech signal selectively over time and frequency in order to reestablish a certain level difference between the average speech spectrum and the measured noise spectrum, i. e., to recover a target signal-to-noise ratio (SNR).

In [2], we presented an enhancement system that uses a non-uniform low delay filter-bank and employs a similar weighting rule as [1]. The processing is performed by means of the frequency warped filter-bank equalizer (FBE), which performs time-domain filtering with coefficients adapted in the frequency domain. This allows for a processing with approximately Bark-scaled spectral resolution and low signal delay.

Shin et al. proposed in [3] the reinforcement of the 'perceptual loudness' of the speech signal to the same level as it would have had in silence. This approach aims primarily at equal loudness, unaltered tone color, and overall quality, which includes among others intelligibility, clarity, naturalness and pleasantness.

Opposed to that, in this contribution we try to improve primarily speech intelligibility and explicitly accept changes in tone color or loudness. We analyze theoretically the influence of the speech spectrum level of each subband on the Speech Intelligibility Index (SII) for a given noise spectrum level. Using the results of this analysis, an improved near end listening enhancement algorithm is proposed which maximizes the SII and thus speech intelligibility. Furthermore, we investigate the impact of ignoring masking effects on the algorithm.

2. SPEECH INTELLIGIBILITY INDEX

The Speech Intelligibility Index (SII) [4] is a standardized objective measure which is correlated with the intelligibility of speech under a variety of adverse listening conditions. In this section the calculation rules of the *critical band procedure* of the SII are analyzed in order to design an improved near end listening enhancement algorithm.

2.1 Calculation Rules of SII

The SII is based on the equivalent speech spectrum level[1] E_i' as well as the equivalent noise spectrum level N_i' in each contributing subband i, both measured in dB. The spectrum level is basically the power average over time in each subband differentiated with respect to the bandwidth of the subband. It can be approximated by the power average over time in each subband divided by its bandwidth [4].

For the application of near end listening enhancement, only those situations with significant background noise are of interest. Therefore, it is feasible to make the following assumptions, which simplify the calculation of the SII:

- We assume that the equivalent noise spectrum level N_i' is greater than the self-speech masking spectrum level $V_i = E_i' - 24\,\text{dB}$ [4], which accounts for the masking of higher speech frequencies by lower speech frequencies. This approximation (if relevant at all) has influence just on the spread of masking.
- We further assume the equivalent masking spectrum level Z_i to be greater than the equivalent internal noise spectrum level [4], which corresponds to the threshold of hearing.

Considering these approximations, the following steps have to be performed for each contributing subband i to calculate the SII:

[1] The equivalent spectrum level is defined as the spectrum level measured at the point corresponding to the center of the listener's head, with the listener absent, under the reference communication situation [4].

1. Determine the slope per octave C_i of the spread of masking caused by the background noise:

$$C_i = -80\,\text{dB} + 0.6\left[N'_i + 10\log(h_i - l_i)\right] \tag{1}$$

with h_i and l_i being the upper and lower limiting frequency of the i-th critical band.

2. Determine the equivalent disturbance spectrum level D_i, which is equal to the equivalent masking spectrum level Z_i due to the assumption made above:

$$D_i = Z_i = 10\log\left\{10^{0.1N'_i} + \sum_{\lambda=1}^{i-1} 10^{0.1\left[N'_\lambda + 3.32C_\lambda \log\left(\frac{f_i}{f_\lambda}\right)\right]}\right\}, \tag{2}$$

where f_i is the center frequency of the i-th critical band.

3. Determine the speech level distortion factor $L_i(E'_i)$:

$$L_i(E'_i) = \begin{cases} 1 & \text{if } E'_i \le U_i + 10\,\text{dB} \\ 1 - \frac{E'_i - U_i - 10\,\text{dB}}{160\,\text{dB}} & \text{if } U_i + 10\,\text{dB} < E'_i < U_i + 170\,\text{dB} \\ 0 & \text{if } U_i + 170\,\text{dB} \le E'_i\,, \end{cases} \tag{3}$$

which allows for the decrease in intelligibility caused by the distortion due to a high presentation level. U_i denotes the standard speech spectrum level at normal voice effort [4, Table 1], which has its maximum value of 34.75 dB in the second critical band with $f_2 = 250\,\text{Hz}$.

4. Determine the band audibility function $A_i(E'_i)$

$$A_i(E'_i) = L_i(E'_i) \cdot K_i(E'_i) \tag{4}$$

using the auxiliary variable[2] $K_i(E'_i)$

$$K_i(E'_i) = \begin{cases} 0 & \text{if } E'_i \le D_i - 15\,\text{dB} \\ \frac{E'_i - D_i + 15\,\text{dB}}{30\,\text{dB}} & \text{if } D_i - 15\,\text{dB} < E'_i \le D_i + 15\,\text{dB} \\ 1 & \text{if } D_i + 15\,\text{dB} < E'_i\,. \end{cases} \tag{5}$$

The auxiliary variable $K_i(E'_i)$ accounts for the loss of intelligibility due to the fact that the speech signal is masked, e. g., by noise, and the band audibility function $A_i(E'_i)$ specifies the effective proportion of the speech dynamic range within the subband that contributes to speech intelligibility. Note, that the dependency of $K_i(E'_i)$ on D_i is omitted for the sake of brevity.

Finally, the Speech Intelligibility Index S is calculated as

$$S = \sum_{i=1}^{i_\text{max}} I_i \cdot A_i(E'_i) \tag{6}$$

using the band importance function I_i [4, Table 1], which characterized the relative significance of the subband to speech intelligibility. Since

$$\sum_{i=1}^{i_\text{max}} I_i = 1\,, \tag{7}$$

the SII can take values from zero to one. Communication systems with $S \ge 0.75$ are considered to be good, those with $S \le 0.45$ poor.

[2] In [4], K_i is called 'temporary variable.'

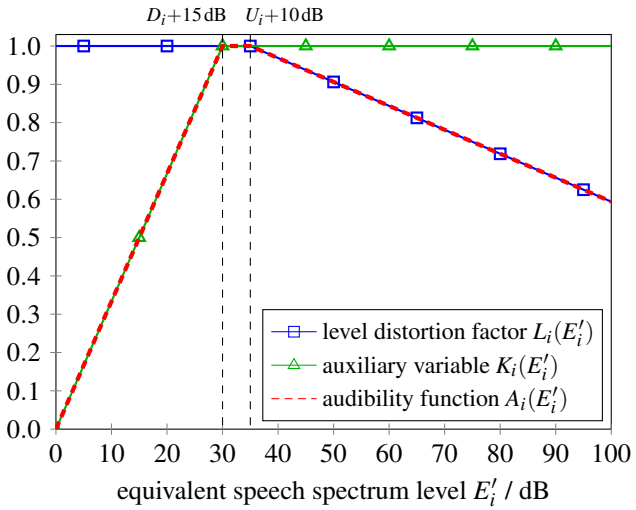

Figure 1: Exemplary plot for case 1 as of Section 2.2.1; $i = 8$, $U_i = 25.01\,\text{dB}$, $D_i = 15\,\text{dB}$.

2.2 Interpretation

The band audibility function $A_i(E'_i)$ as a function of E'_i is determined by two factors with diametrically opposed impact:

- The auxiliary variable $K_i(E'_i)$ *increases* monotonically with increasing equivalent speech spectrum level E'_i.
- The level distortion factor $L_i(E'_i)$ *decreases* monotonically with increasing equivalent speech spectrum level E'_i.

Both functions of E'_i are piecewise linear as defined in (3) and (5). As a consequence, three cases exist depending on the equivalent disturbance spectrum level D_i, which are discussed in the following:

1. The segment with increasing $K_i(E'_i)$ ends before the start of the segment with decreasing $L_i(E'_i)$,
2. the segments with increasing $K_i(E'_i)$ and with decreasing $L_i(E'_i)$ overlap, and
3. the segment with increasing $K_i(E'_i)$ starts after $L_i(E'_i)$ has decreased completely. This case is not of practical interest since it occurs only for $D_i > U_i + 185\,\text{dB}$.

2.2.1 *Case 1: $D_i \le U_i - 5\,\text{dB}$*

An example for this case is sketched in Figure 1. With increasing equivalent speech spectrum level E'_i, the auxiliary variable $K_i(E'_i)$ reaches its maximum before the level distortion factor $L_i(E'_i)$ starts to decrease. The resulting band audibility function $A_i(E'_i)$ is continuous and piecewise linear:

$$A_i(E'_i) = \begin{cases} 0 & \text{if } E'_i \le D_i - 15\,\text{dB} \\ \frac{E'_i - D_i + 15\,\text{dB}}{30\,\text{dB}} & \text{if } D_i - 15\,\text{dB} < E'_i \le D_i + 15\,\text{dB} \\ 1 & \text{if } D_i + 15\,\text{dB} < E'_i \le U_i + 10\,\text{dB} \\ 1 - \frac{E'_i - U_i - 10\,\text{dB}}{160\,\text{dB}} & \text{if } U_i + 10\,\text{dB} < E'_i \le U_i + 170\,\text{dB} \\ 0 & \text{if } U_i + 170\,\text{dB} < E'_i\,. \end{cases} \tag{8}$$

It can be seen, that the maximum value

$$\max_{\tilde{E}'_i} A_i(\tilde{E}'_i) = 1 \tag{9}$$

is reached for

$$D_i + 15\,\text{dB} \le E'_i \le U_i + 10\,\text{dB}\,. \tag{10}$$

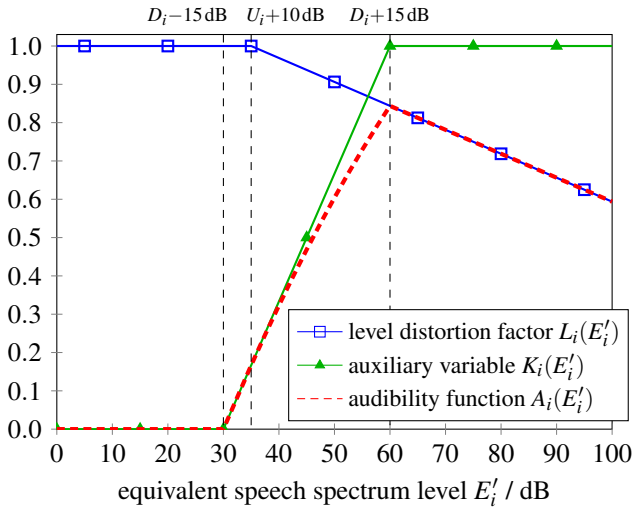

Figure 2: Exemplary plot for case 2 as of Section 2.2.2; $i = 8$, $U_i = 25.01\,\mathrm{dB}$, $D_i = 45\,\mathrm{dB}$.

2.2.2 *Case 2: $D_i > U_i - 5\,\mathrm{dB}$*

An example for this case is sketched in Figure 2. The increasing segment of the auxiliary variable $K_i(E_i')$ overlaps with the decreasing segment of the level distortion factor $L_i(E_i')$. The resulting band audibility function $A_i(E_i')$ is continuous, a downward opened parabola in the overlapping segment and piecewise linear elsewhere:

$$A_i(E_i') = \begin{cases} 0 & \text{if } E_i' \le D_i - 15\,\mathrm{dB} \\ \frac{E_i' - D_i + 15\,\mathrm{dB}}{30\,\mathrm{dB}} & \text{if } D_i - 15\,\mathrm{dB} < E_i' \le \zeta \\ \left(\frac{E_i' - D_i + 15\,\mathrm{dB}}{30\,\mathrm{dB}}\right) \cdot \left(1 - \frac{E_i' - U_i - 10\,\mathrm{dB}}{160\,\mathrm{dB}}\right) & \text{if } \zeta < E_i' \le \xi \\ 1 - \frac{E_i' - U_i - 10\,\mathrm{dB}}{160\,\mathrm{dB}} & \text{if } \xi < E_i' \le U_i + 170\,\mathrm{dB} \\ 0 & \text{if } U_i + 170\,\mathrm{dB} < E_i' \end{cases} \tag{11}$$

with $\zeta = \max\{U_i + 10\,\mathrm{dB},\, D_i - 15\,\mathrm{dB}\}$
and $\xi = \min\{D_i + 15\,\mathrm{dB},\, U_i + 170\,\mathrm{dB}\}$.

In the third segment ($\zeta < E_i' \le \xi$), the gradient

$$\frac{\mathrm{d}A_i(E_i')}{\mathrm{d}E_i'} = \frac{1}{30\,\mathrm{dB}} \cdot \left(1 - \frac{E_i' - U_i - 10\,\mathrm{dB}}{160\,\mathrm{dB}}\right) - \frac{E_i' - D_i + 15\,\mathrm{dB}}{30\,\mathrm{dB}} \cdot \frac{1}{160\,\mathrm{dB}} \tag{12}$$

is always positive if $D_i < U_i + 125\,\mathrm{dB}$. Thus

$$\frac{\mathrm{d}A_i(E_i')}{\mathrm{d}E_i'} \begin{cases} = 0 & \text{if } E_i' \le D_i - 15\,\mathrm{dB} \\ > 0 & \text{if } D_i - 15\,\mathrm{dB} < E_i' \le \zeta \\ > 0 & \text{if } \zeta < E_i' \le \xi \\ < 0 & \text{if } \xi < E_i' \le U_i + 170\,\mathrm{dB} \\ = 0 & \text{if } U_i + 170\,\mathrm{dB} < E_i'\,. \end{cases} \tag{13}$$

It follows, that the maximum value

$$\max_{\tilde{E}_i'} A_i(\tilde{E}_i') = 1 - \frac{D_i - U_i + 5\,\mathrm{dB}}{160\,\mathrm{dB}} \tag{14}$$

is reached if

$$E_i' = \xi = D_i + 15\,\mathrm{dB}\,. \tag{15}$$

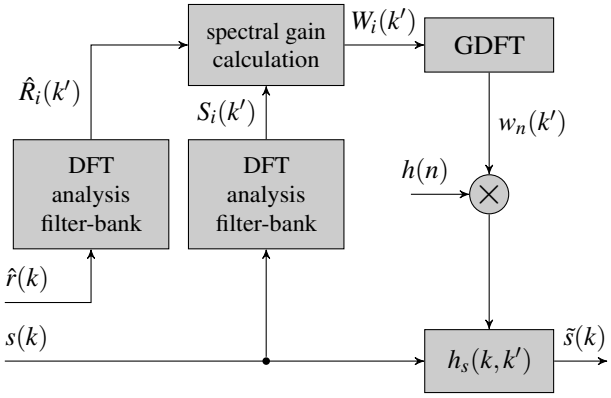

Figure 3: System for near end listening enhancement.

2.2.3 *Conclusions*

For the practically relevant case $D_i < U_i + 125\,\mathrm{dB}$, it follows for all equivalent disturbance spectrum levels D_i that the smallest equivalent speech spectrum level E_i' which results in the maximum value of the band audibility function $A_i(E_i')$

$$\max_{\tilde{E}_i'} A_i(\tilde{E}_i') = \min\left\{1,\; 1 - \frac{D_i - U_i + 5\,\mathrm{dB}}{160\,\mathrm{dB}}\right\} \tag{16}$$

is given by

$$\min\left\{E_i' \mid A_i(E_i') = \max_{\tilde{E}_i'} A_i(\tilde{E}_i')\right\} = D_i + 15\,\mathrm{dB}\,. \tag{17}$$

Accordingly, given the equivalent disturbance spectrum level D_i, the theoretically maximum SII

$$\max_{\substack{E_i' \\ 1 \le i \le i_{\max}}} S = \sum_{i=1}^{i_{\max}} I_i \cdot \min\left\{1,\; 1 - \frac{D_i - U_i + 5\,\mathrm{dB}}{160\,\mathrm{dB}}\right\} \tag{18}$$

is acchieved for the equivalent speech spectrum level

$$E_i' = D_i + 15\,\mathrm{dB} \tag{19}$$

in each subband i.

3. NEAR END LISTENING ENHANCEMENT

We utilize the system for near end listening enhancement by means of the warped filter-bank equalizer (FBE) as described in [2] and depicted in Figure 3. Opposed to the discrete Fourier transform (DFT) analysis-synthesis filter-bank, which is conventionally used for speech enhancement, this structure allows for a processing with approximately Bark-scaled spectral resolution and low signal delay.

The (clean) far end speech signal $s(k)$ and the near end noise estimate $\hat{r}(k)$ are split into M subband signals $S_i(k')$ and $\hat{R}_i(k')$ by means of a warped DFT analysis filter-bank with downsampling. The subsampled time index is given by $k' = \lfloor k/R' \rfloor \cdot R'$ where R' marks the downsampling rate. The real impulse response of the prototype filter of length $L+1$ is denoted by $h(n)$.

The non-uniform time-frequency resolution is designed by means of an allpass transformation, which achieves a variation of the subband filter bandwidths without changing filter

properties such as stopband attenuation etc. An allpass pole of $a = 0.4$ yields a good approximation of the Bark frequency scale for the considered sampling rate of $f_s = 8\,\text{kHz}$, cf. [5].

The subband signals $S_i(k')$ and $\hat{R}_i(k')$ are used to calculate the spectral gains $W_i(k')$ as described later in Section 3.1. The enhanced speech signal $\tilde{s}(k)$ is obtained by filtering the far end speech signal $s(k)$ with time-varying filter coefficients, which are obtained by a generalized discrete Fourier transform (GDFT) of the spectral weights $W_i(k')$.

It should be noted that only a rough overview of the FBE and the utilized enhancement system is given here. These aspects are treated in [6, 7] and [2] in more detail.

3.1 SII Optimized Algorithm

In order to calculate the time-varying gain factors $W_i(k')$, the short-term power spectral densities (PSDs) $\Phi_{ss,i}(k')$ and $\Phi_{rr,i}(k')$ of the subband signals $S_i(k')$ and $R_i(k')$ are transformed to the equivalent spectrum levels $E'_i(k')$ and $N'_i(k')$ as described in the first paragraph of Section 2.1:

$$E'_i(k') = 10\log\left\{\frac{g_l^2 \cdot \Phi_{ss,i}(k')}{h_i - l_i}\right\}, \tag{20}$$

$$N'_i(k') = 10\log\left\{\frac{g_l^2 \cdot \Phi_{rr,i}(k')}{h_i - l_i}\right\}, \tag{21}$$

where h_i and l_i are the upper and lower limiting frequency of the i-th critical band and g_l is a normalization factor to make the analysis filter-bank of the FBE approximately 'lossless':

$$g_l = \frac{1}{\sqrt{M \cdot \sum_{n=0}^{M-1} h^2(n)}}. \tag{22}$$

The equivalent speech spectrum level $\tilde{E}'_i(k')$ of the amplified speech $\tilde{S}_i(k') = W'_i(k') \cdot S_i(k')$ can be calculated in analogy to (20) as

$$\tilde{E}'_i(k') = 10\log\left\{\frac{g_l^2 \cdot \Phi_{\tilde{s}\tilde{s},i}(k')}{h_i - l_i}\right\} \tag{23}$$

$$= 10\log\left\{\frac{g_l^2 \cdot W'_i(k')^2 \cdot \Phi_{ss,i}(k')}{h_i - l_i}\right\} \tag{24}$$

$$= 20\log\left\{W'_i(k')\right\} + E'_i(k'). \tag{25}$$

Next, the equivalent disturbance spectrum level $D_i(k')$ is calculated according to (1) and (2). Finally, based on (19), the time-varying gain factors $W'_i(k')$ are chosen such that

$$\tilde{E}'_i(k') = D_i(k') + 15\,\text{dB}. \tag{26}$$

Furthermore, the speech signal should not be attenuated in a noise-free environment, which leads with (25) to the gain

$$W'_i(k') = \max\left\{10^{0.05\left[D_i(k') + 15\,\text{dB} - E'_i(k')\right]}, 1\right\}. \tag{27}$$

In order to prevent pain and hearing damage, the gain is limited such that the resulting instantaneous equivalent spectrum level of the amplified speech in each subband does not exceed a maximum spectrum level $E'_{\max} = 90\,\text{dB}$:

$$W_i(k') = \min\left\{W'_i(k'), \sqrt{\frac{10^{0.1E'_{\max}}}{g_l^2 \cdot |S_i(k')|^2}}\right\}. \tag{28}$$

The value of $E'_{\max}$ is chosen in accordance to [8, Fig. 2.1].

3.2 SNR Recovery Algorithm

If the spread of masking is neglected, i. e., the slope per octave of the spread of masking is approximated as

$$C_i(k') = -\infty\,\text{dB}, \tag{29}$$

the equivalent disturbance spectrum level of (2) reduces to

$$D_i(k') = N'_i(k'). \tag{30}$$

Using (20), (21), and (27) the time-varying gain factors turn out to be

$$W'_i(k') = \max\left\{\sqrt{\xi \frac{\Phi_{rr,i}(k')}{\Phi_{ss,i}(k')}}, 1\right\} \tag{31}$$

with $\xi \mathrel{\widehat{=}} 15\,\text{dB}$. This is exactly the SNR recovery algorithm which was found heuristically in [2].

Our previous solutions for near end listening enhancement as proposed in [1] and [2] limit the gain factors to a fixed maximum gain $W_{\max}$ of, e. g., $W_{\max} \mathrel{\widehat{=}} 30\,\text{dB}$. This was done to prevent 'over-amplification' of narrow subband or single spectral components with low energy, since this would interfere with the spectral fine structure such as harmonics of the speech signal. If spectral weighting is performed in subbands as wide as critical bands, this interference is reduced and the limitation to a fixed maximum gain becomes obsolete.

Instead of a fixed maximum gain, we propose to use the same limiting of the resulting instantaneous equivalent spectrum level of the amplified speech in each subband as in (28).

4. RESULTS

The performance of the two proposed algorithms was evaluated in terms of the SII using the so-called critical band procedure [4] for every speech file of the TIMIT database, in total 5.4 hours, disturbed by the *factory1* noise from the NOISEX-92 database at a sampling rate of 8 kHz.

In order to calculate the speech and noise spectrum level of each sound file, the spectrum level is averaged for half-overlapping Hann-windowed frames of 20 ms length. Finally, the average SII over all speech files is taken. As noted above, good communication systems have an SII of 0.75 or better while the SII of poor communication systems is below 0.45.

Prior to processing, the speech database is scaled to match the overall sound pressure level of 62.35 dB as specified in [4] for normal voice effort. The desired input SNR is achieved by adjusting the sound pressure level of the noise file in relation to 62.35 dB.

In Figure 4 the average Speech Intelligibility Index is plotted after processing with

- the proposed SII optimized algorithm of Section 3.1,
- the modified SNR recovery algorithm of Section 3.2,
- the SNR recovery algorithm as described in [2], and
- without processing.

As an additional reference, the established theoretical bound for the maximum achievable SII from (18) is also depicted.

It can be seen, that all three above mentioned algorithms have almost the same performance for SNRs above $-10\,\text{dB}$. For worse SNRs, the performance of the SNR recovery algorithm with fixed maximum gain deteriorates rapidly due to the strong limiting.

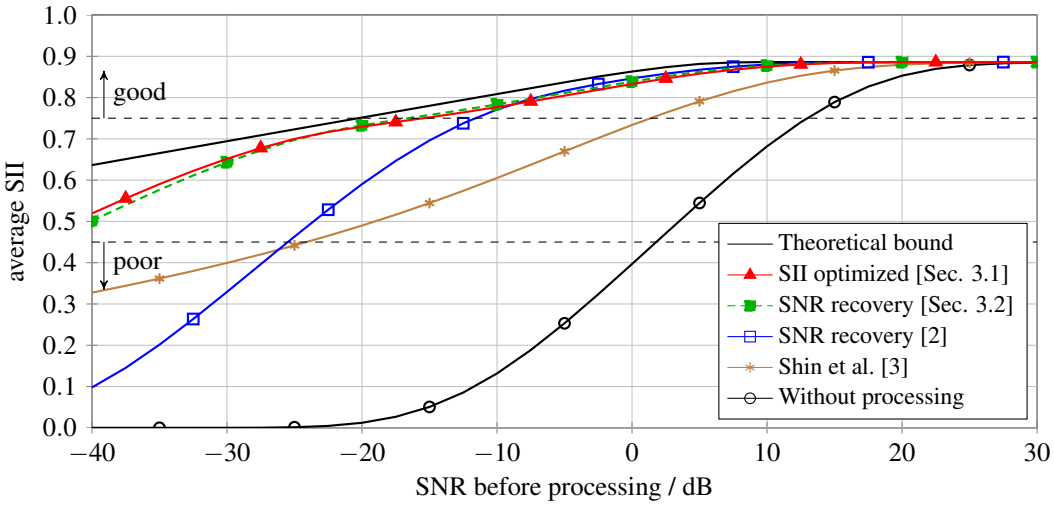

Figure 4: Comparison of average SII after processing with proposed algorithms as well as with algorithm of Shin et al. with theoretical bound and average SII without processing as reference.

Note, that the algorithms optimize the SII for each frame separately based on the current (smoothed) spectrum levels, whereas the average SII is calculated from the mean spectrum level over each entire sound file. Due to that, the SNR recovery algorithms perform very slightly better than the SII optimized algorithm in terms of average SII for input SNRs above −20 dB although it neglects the masking effect and hence leads to an equal or smaller SII in each frame.

Nevertheless, the slope of masking apparently has only insignificant influence on the performance of the algorithm. The listening experience supports this finding.

Both proposed algorithms perform close to the theoretical bound unless they are limited to prevent hearing damage. The remaining difference occurs for the reasons mentioned above.

The proposed algorithms are also compared to the algorithm of Shin et al. [3], which performs worse in terms of SII for the considered input SNRs. This is due to the fact that Shin et al. primarily aim at unaltered tone color whereas the SII optimized algorithm tries to improve speech intelligibility without respect to tone color.

Sound samples and further information can be found at `http://www.ind.rwth-aachen.de/~bib/sauert09/`.

5. CONCLUSIONS

In this contribution, the influence of the equivalent speech spectrum level of each subband on the Speech Intelligibility Index (SII) given an equivalent noise spectrum level is theoretically analyzed. The major result is that the SII is maximized for all practically relevant equivalent noise spectrum levels if the equivalent speech spectrum level is 15 dB above the equivalent noise spectrum level in each subband.

These findings are used to derive a new SII optimized near end listening enhancement algorithm. Furthermore, the disregard of masking effects during the derivation of the SII optimized algorithm directly leads to the SNR recovery algorithm as proposed in [2] and also gives a theoretical explanation of the heuristical target SNR used there.

The instrumental evaluation by means of the SII has shown that the new algorithm performs close to the theoritical bound for the maximum achievable SII given the equivalent noise spectrum level. Since the SNR recovery algorithm performs very similar, it further shows that accounting for masking effects in this algorithm has only slight influence on speech intelligibility.

REFERENCES

[1] Bastian Sauert and Peter Vary. Near end listening enhancement: Speech intelligibility improvement in noisy environments. In *Proc. of Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 493–496, Toulouse, France, May 2006.

[2] Bastian Sauert, Heinrich W. Löllmann, and Peter Vary. Near end listening enhancement by means of warped low delay filter-banks. In *Proc. of ITG-Fachtagung Sprachkommunikation*, Aachen, Germany, October 2008.

[3] Jong Won Shin, Woohyung Lim, Junesig Sung, and Nam Soo Kim. Speech reinforcement based on partial specific loudness. In *Proc. of European Conf. on Speech Communication and Technology (EUROSPEECH)*, pages 978–981, Antwerp, Belgium, August 2007.

[4] American National Standard. Methods for the Calculation of the Speech Intelligibility Index. ANSI S3.5-1997, 1997.

[5] Julius O. Smith, III and Jonathan S. Abel. Bark and erb bilinear transforms. *IEEE Transactions on Speech and Audio Processing*, 7(6):697–708, November 1999.

[6] Heinrich W. Löllmann and Peter Vary. Uniform and warped low delay filter-banks for speech enhancement. *Speech Communication*, 49:574–587, July 2007. Special issue on Speech Enhancement.

[7] Heinrich W. Löllmann and Peter Vary. Low delay filter-banks for speech and audio processing. In E. Hänsler and G. Schmidt, editors, *Speech and Audio Processing in Adverse Environments*, chapter 2, pages 13–61. Springer, Berlin, New York, 2008.

[8] Eberhard Zwicker and Hugo Fastl. *Psychoacoustics. Facts and Models*. Springer, Berlin, Heidelberg, New York, 2nd edition, 1999.