NEAR END LISTENING ENHANCEMENT OPTIMIZED WITH RESPECT TO SPEECH INTELLIGIBILITY INDEX AND AUDIO POWER LIMITATIONS

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

In speech communications, 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 telephony in acoustical background noise such as traffic or babble noise. In these situations, it is often not acceptable/possible to increase the audio power amplification.

In this contribution we use a theoretical analysis of the Speech Intelligibility Index (SII) to develop an algorithm which numerically maximizes the SII under the constraint of an unchanged average power of the audio signal.

1. INTRODUCTION

Mobile telephony 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* (downlink) speech and the acoustical background noise from the *near end* and thus experiences a reduced speech intelligibility.

As the noise signal cannot be influenced, a reasonable approach to improve intelligibility by near end listening enhancement is to manipulate the *far end* speech signal in dependence of the *near end* background noise as shown in Figure 1.

Several approaches for near end listening enhancement are known from literature, e. g., [1, 2, 3, 4, 5, 6].

In [3], we presented a theoretical analysis of the influence of the speech spectrum level on the Speech Intelligibility Index (SII) for a given noise spectrum level. Using this analysis, an improved near end listening enhancement algorithm was derived which maximizes the SII and thus speech intelligibility by frequency selective raising of the speech signal power. The processing is performed by means of a 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.

However, in some applications the power of the loudspeaker signal is constrained to the power of the original signal, e. g., because the sound reproduction system has no head-room in terms of output audio power. This applies, e. g., for mobile phones with tiny loudspeakers which are already saturated at low output power levels.

In this contribution we use a theoretical analysis of the SII to maximize the SII numerically under the constraint of an unchanged average power of the audio signal.

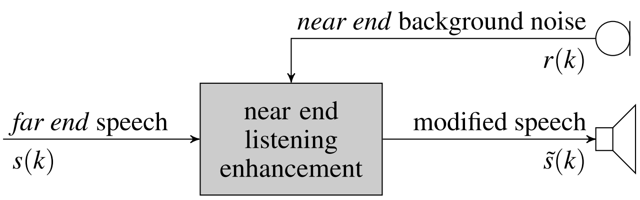


Figure 1: Principle of near end listening enhancement.

2. SPEECH INTELLIGIBILITY INDEX

The Speech Intelligibility Index (SII) [7] is a standardized objective measure which is correlated with the intelligibility of speech under a variety of adverse listening conditions.

2.1 Calculation Rules of SII

In this section the calculation rules of the *critical band procedure* of the SII are briefly presented. 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, which can be approximated by the average power in each subband with reference pressure 20 µPa divided by its bandwidth measured in dB [7].

For the application of near end listening enhancement, only 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 so-called self-speech masking spectrum level $V_i = E_i - 24\,\mathrm{dB}$ [7], 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 [7], which corresponds to the threshold of hearing.

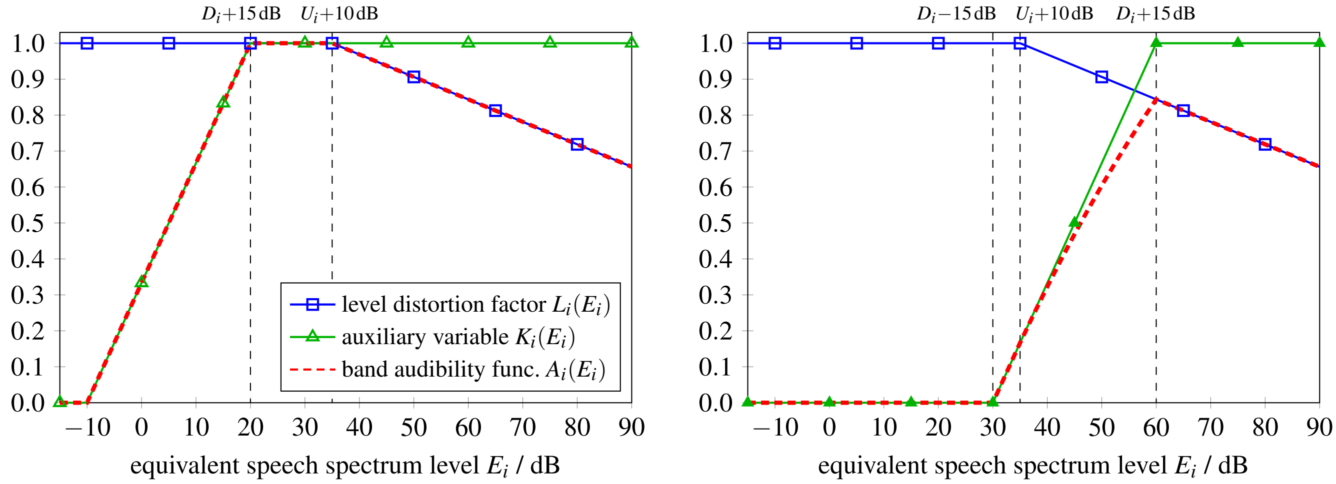
Considering these approximations, the following steps have to be performed for each contributing subband i to calculate the SII:

1. 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:

$$D_i = Z_i = 10\log\left\{10^{N_i/10} + \sum_{\lambda=1}^{i-1} 10^{[N_\lambda + 3.32 C_\lambda \log(f_i/h_\lambda)]/10}\right\}, \tag{1}$$

[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 [7].
It is denoted by E_i in this paper as opposed to E_i' in [7] for the sake of clarity.

 ISSN 2076-1465



(a) Disjunct case: $i = 8$, $U_i = 25.01\,\text{dB}$, $D_i = 5\,\text{dB}$.

(b) Overlapping case: $i = 8$, $U_i = 25.01\,\text{dB}$, $D_i = 45\,\text{dB}$.

Figure 2: Contributions to the band audibility function for the cases of Section 2.2.

where

$$C_i = -80\,\text{dB} + 0.6\big[N_i + 10\log(h_i - l_i)\big] \tag{2}$$

is the slope per octave of the spread of masking caused by the background noise. h_i, l_i, and f_i denote the upper and lower limiting and center frequencies of the i-th subband respectively.

2. 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 considers the decrease in intelligibility caused by the distortion due to a high presentation level. U_i is fixed and denotes the standard speech spectrum level at normal voice effort [7, Table 1], which has its maximum value of 34.75 dB in the second critical band with $f_2 = 250\,\text{Hz}$.

3. 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 ('temporary') variable $K_i(E_i)$ with

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

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. 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.

Finally, the Speech Intelligibility Index S is calculated as

$$S = \sum_{i=1}^{i_{\max}} I_i \cdot A_i(E_i) \tag{6}$$

with the number of subbands $i_{\max}$. The band importance function I_i [7, Table 1] characterizes the relative significance of the subband to speech intelligibility. Since $\sum_{i=1}^{i_{\max}} I_i = 1$ and $0 \le A_i \le 1$, the SII can take values from zero to one.

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 for (4) depending on the equivalent disturbance spectrum level D_i:

1. *Disjunct case:* The segment with increasing $K_i(E_i)$ ends before the start of the segment with decreasing $L_i(E_i)$. An example for this case is sketched in Figure 2a.
2. *Overlapping case:* The segments with increasing $K_i(E_i)$ and with decreasing $L_i(E_i)$ overlap, which is exemplified in Figure 2b.
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 $D_i > U_i + 185\,\text{dB}$.

3. NEAR END LISTENING ENHANCEMENT

In this section, we give an overview of our system for near end listening enhancement by means of a warped filter-bank equalizer (FBE) as depicted in Figure 3. The details are treated in [8] and [9, 10]. In contrast to the discrete Fourier transform (DFT) analysis-synthesis filter-bank, which is widely used for speech enhancement, this structure allows for an efficient processing with approximately Bark-scaled spectral resolution and low signal delay.

The (clean) far end speech signal $s(k)$ and the near end noise $r(k)$ are split into M subband signals $S_i(k')$ and $R_i(k')$ by means of a warped DFT analysis filter-bank with downsampling. The time index in the subsampled domain is given by k'. The real-valued impulse response of the prototype lowpass filter has length $L+1$. In this paper, L is chosen equal to M.

The non-uniform time-frequency resolution is achieved by means of an allpass transformation, which accomplishes a variation of the subband filter bandwidths without changing certain filter properties such as stopband attenuation. An allpass pole of $a = 0.4$ and a DFT size of $M = 34$ yield a good

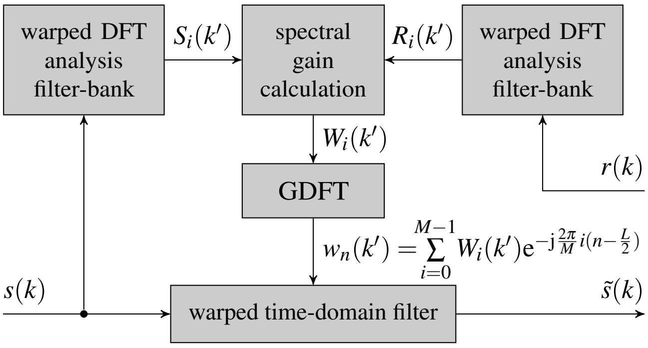


Figure 3: System for near end listening enhancement with time index k, subsampled time index k', subband index i, and $0 \leq n \leq L$.

approximation of the Bark frequency scale for the considered sampling rate of $f_s = 8\,\mathrm{kHz}$. Since the first SII subband begins at 100 Hz, this results in $i_{\max} = 17$ non-redundant SII subbands.

The subband signals $S_i(k')$ and $R_i(k')$ are used to calculate the spectral gains $W_i(k')$ as described later in Section 3.5. 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 $W_i(k')$.

3.1 Calculation of Spectrum Levels

The equivalent spectrum levels $E_i(k')$ and $N_i(k')$ are computed as described in the first paragraph of Section 2.1:

$$E_i(k') = 10\log\left\{\frac{g_l^2 \cdot \Phi_{ss,i}(k')}{\Delta f_i}\right\}, \tag{7}$$

$$N_i(k') = 10\log\left\{\frac{g_l^2 \cdot \Phi_{rr,i}(k')}{\Delta f_i}\right\}, \tag{8}$$

where $\Delta f_i = h_i - l_i$ is the frequency bandwidth of the i-th critical band. The short-term power spectral densities $\Phi_{ss,i}(k')$ and $\Phi_{rr,i}(k')$ are determined as the recursively smoothed squared norm of the subband signals $S_i(k')$ and $R_i(k')$ as described in [8]. g_l is a normalization factor to achieve approximately a unity gain analysis filter-bank of the FBE:

$$g_l = \left(\sqrt{M \cdot \sum_{n=0}^{L} h^2(n)}\right)^{-1}. \tag{9}$$

The equivalent speech spectrum level $\tilde{E}_i(k')$ of the modified speech signal $\tilde{S}_i(k') = W_i'(k') \cdot S_i(k')$ can be calculated in analogy to (7) as

$$\tilde{E}_i(k') = 10\log\left\{\frac{g_l^2 \cdot W_i'(k')^2 \cdot \Phi_{ss,i}(k')}{\Delta f_i}\right\} \tag{10}$$

$$= 20\log\left\{W_i'(k')\right\} + E_i(k'). \tag{11}$$

Finally, the equivalent disturbance spectrum level $D_i(k')$ is calculated according to (1).

3.2 Audio Power Limitation

As noted above, the SII should be maximized under the constraint that the short-term audio power of the optimized output signal is less or equal than the short-term audio power $P_{\mathrm{ref}}(k')$ of the input signal:

$$\sum_{i=1}^{i_{\max}} \Delta f_i \cdot 10^{\tilde{E}_i(k')/10} \overset{!}{\leq} \sum_{i=1}^{i_{\max}} \Delta f_i \cdot 10^{E_i(k')/10} =: P_{\mathrm{ref}}(k'). \tag{12}$$

In the following two sections, the dependency on k' is not written down for simplicity.

3.3 Approximation by Concave Function

With regard to the envisaged iterative numerical optimization scheme, the band audibility function is approximated by a strictly concave function $\hat{A}_i(\tilde{E}_i)$ to ensure convergence to a global maximum. For this purpose, the limitations to zero of the level distortion factor $L_i(\tilde{E}_i)$ (third case of (3)) and of the auxiliary variable $K_i(\tilde{E}_i)$ (first case of (5)) are omitted.

In the disjunct case (see also Figure 4), the constant segment between $D_i + 15\,\mathrm{dB}$ and $U_i + 10\,\mathrm{dB}$ is replaced by a very slight linear ascent for $\tilde{E}_i < E_i$ and a linear descent for $\tilde{E}_i > E_i$. This way, the function becomes strictly concave and the gain factor in (11) tends towards one if no further improvement in SII can be achieved, which reduces speech distortion.

This finally results in the disjunct case ($D_i + 15\,\mathrm{dB} < U_i + 10\,\mathrm{dB}$) in

$$\hat{A}_i(\tilde{E}_i) = \begin{cases} \frac{\tilde{E}_i - D_i + 15\,\mathrm{dB}}{30\,\mathrm{dB}} & \text{if } \tilde{E}_i \leq D_i + 15\,\mathrm{dB} \\ \hat{A}_i(D_i{+}15\,\mathrm{dB}) + \frac{\tilde{E}_i - D_i - 15\,\mathrm{dB}}{300\,\mathrm{dB}} & \text{if } D_i + 15\,\mathrm{dB} < \tilde{E}_i \leq E_i^* \\ \hat{A}_i(E_i^*) - \frac{\tilde{E}_i - E_i^*}{300\,\mathrm{dB}} & \text{if } E_i^* < \tilde{E}_i \leq U_i + 10\,\mathrm{dB} \\ \hat{A}_i(U_i{+}10\,\mathrm{dB}) - \frac{\tilde{E}_i - U_i - 10\,\mathrm{dB}}{160\,\mathrm{dB}} & \text{if } U_i + 10\,\mathrm{dB} < \tilde{E}_i \end{cases} \tag{13}$$

with $E_i^* = \max\{\min\{E_i, U_i + 10\,\mathrm{dB}\},\ D_i + 15\,\mathrm{dB}\}$

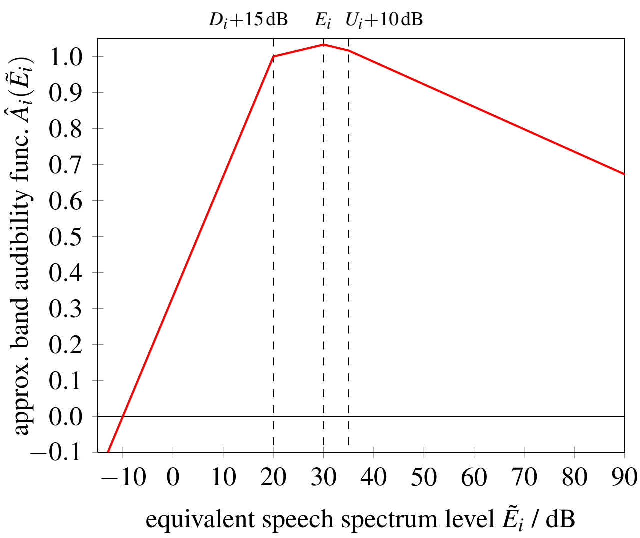


Figure 4: Exemplary plot of concave approximation of band audibility function (compare Figure 2a); $i = 8$, $U_i = 25.01\,\mathrm{dB}$, $D_i = 5\,\mathrm{dB}$, $E_i = 30\,\mathrm{dB}$.

and in the overlapping case ($D_i + 15\,\mathrm{dB} \geq U_i + 10\,\mathrm{dB}$) in

$$\hat{A}_i(\tilde{E}_i) = \begin{cases} \frac{\tilde{E}_i - D_i + 15\,\mathrm{dB}}{30\,\mathrm{dB}} & \text{if } \tilde{E}_i \leq \zeta \\ \left(\frac{\tilde{E}_i - D_i + 15\,\mathrm{dB}}{30\,\mathrm{dB}}\right) \cdot \left(1 - \frac{\tilde{E}_i - U_i - 10\,\mathrm{dB}}{160\,\mathrm{dB}}\right) & \\ & \text{if } \zeta < \tilde{E}_i \leq D_i + 15\,\mathrm{dB} \\ 1 - \frac{\tilde{E}_i - U_i - 10\,\mathrm{dB}}{160\,\mathrm{dB}} & \text{if } D_i + 15\,\mathrm{dB} < \tilde{E}_i \end{cases} \quad (14)$$

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

It follows from (13) as well as (14) with the definition of E_i^* and the condition for the overlapping case, that the maximum SII would be achieved for $\tilde{E}_i = E_i^*$ in each subband independent of the case. Hence, if E_i^* fulfills the audio power constraint (12), the optimum solution $\tilde{E}_i^{\mathrm{opt}} = E_i^*$ is found and no further steps are necessary. This will be the case in low-noise conditions or if the equivalent speech spectrum of the input speech signal is higher than $U_i + 10\,\mathrm{dB}$.

3.4 Numerical Optimization

If, on the other hand, E_i^* does not fulfill the audio power constraint (12), the optimum equivalent speech spectrum level $\tilde{E}_i^{\mathrm{opt}}$ must fulfill the equality condition in (12). In this case, the equality constraint optimization problem is transformed into a bounded optimization problem by expressing $\tilde{E}_1$ as a function of $\underline{\tilde{E}} = (\tilde{E}_2, \tilde{E}_3, \ldots, \tilde{E}_{i_{\max}})^{\mathrm{T}}$:

$$\tilde{E}_1(\underline{\tilde{E}}) = 10 \log \left(\frac{P_1(\underline{\tilde{E}})}{\Delta f_1} \right) \quad (15)$$

with

$$P_1(\underline{\tilde{E}}) = P_{\mathrm{ref}} - \sum_{i=2}^{i_{\max}} \Delta f_i \cdot 10^{\tilde{E}_i/10}, \quad (16)$$

leading with (6) to the strictly concave optimization function

$$\hat{S}'(\underline{\tilde{E}}) = I_1 \cdot \hat{A}_1(\tilde{E}_1(\underline{\tilde{E}})) + \sum_{i=2}^{i_{\max}} I_i \cdot \hat{A}_i(\tilde{E}_i). \quad (17)$$

In concordance with [7], the equivalent speech spectrum $\tilde{E}_i$ is bounded with $\tilde{E}_i \geq -50\,\mathrm{dB}$. In the case of $\tilde{E}_1$, this is done with a penalty function for $P_1(\underline{\tilde{E}}) < \Delta f_1 \cdot 10^{-50\,\mathrm{dB}/10} =: P_{1,\min}$. This results in the final function to be maximized:

$$\hat{S}(\underline{\tilde{E}}) = \begin{cases} -100 + P_1(\underline{\tilde{E}}) & \text{if } P_1(\underline{\tilde{E}}) < P_{1,\min} \\ I_1 \cdot \hat{A}_1(\tilde{E}_1(\underline{\tilde{E}})) + \sum_{i=2}^{i_{\max}} I_i \cdot \hat{A}_i(\tilde{E}_i) & \text{otherwise}, \end{cases} \quad (18)$$

which is exemplarily plotted in Figure 5. The solution $\underline{\tilde{E}}^{\mathrm{opt}}(k')$ of this bounded nonlinear multivariable optimization problem is found using the MATLAB function `fmincon` with the trust-region-reflective algorithm. The optimum equivalent speech spectrum level of the preceeding update interval $\underline{\tilde{E}}^{\mathrm{opt}}(k'-1)$ is used as initial estimate for the solution in order to reduce the number of iterations.

In order to use the trust-region-reflective algorithm, the partial first-order derivative must be supplied. For the optimization function of (18), it is defined and continuous over the whole domain besides the boundaries between the cases of (13) and (14). At these places, we define the partial first-order derivative as the left-sided derivative, leading to

$$\frac{\partial \hat{S}(\underline{\tilde{E}})}{\partial \tilde{E}_i} = \begin{cases} -\frac{\ln(10)}{10} \cdot \Delta f_i \cdot 10^{\tilde{E}_i/10} & \text{if } P_1(\underline{\tilde{E}}) < P_{1,\min} \\ I_1 \cdot \frac{\partial \hat{A}_1(\tilde{E}_1)}{\partial \tilde{E}_1} \frac{\partial \tilde{E}_1(\underline{\tilde{E}})}{\partial \tilde{E}_i} + I_i \cdot \frac{\partial \hat{A}_i(\tilde{E}_i)}{\partial \tilde{E}_i} & \text{otherwise} \end{cases} \quad (19)$$

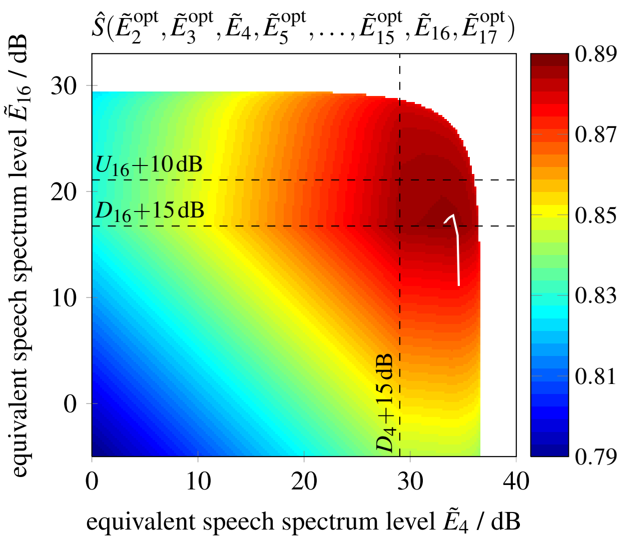


Figure 5: Exemplary plot of optimization function (18) depending on $\tilde{E}_4$ and $\tilde{E}_{16}$; $E_i = U_i$, D_i = average spectrum level of *factory1* noise at SNR of 15 dB; the white line is the optimization trajectory.

with

$$\frac{\partial \tilde{E}_1(\underline{\tilde{E}})}{\partial \tilde{E}_i} = -\frac{\Delta f_i \cdot 10^{\tilde{E}_i/10}}{P_1} \quad (20)$$

as well as

$$\frac{\partial \hat{A}_i(\tilde{E}_i)}{\partial \tilde{E}_i} = \begin{cases} \frac{1}{30\,\mathrm{dB}} & \text{if } \tilde{E}_i \leq D_i + 15\,\mathrm{dB} \\ \frac{1}{300\,\mathrm{dB}} & \text{if } D_i + 15\,\mathrm{dB} < \tilde{E}_i \leq E_i^* \\ -\frac{1}{300\,\mathrm{dB}} & \text{if } E_i^* < \tilde{E}_i \leq U_i + 10\,\mathrm{dB} \\ -\frac{1}{160\,\mathrm{dB}} & \text{if } U_i + 10\,\mathrm{dB} < \tilde{E}_i \end{cases} \quad (21)$$

if $D_i + 15\,\mathrm{dB} < U_i + 10\,\mathrm{dB}$ (disjunct case) and

$$\frac{\partial \hat{A}_i(\tilde{E}_i)}{\partial \tilde{E}_i} = \begin{cases} \frac{1}{30\,\mathrm{dB}} & \text{if } \tilde{E}_i \leq \zeta \\ \frac{-2\tilde{E}_i + U_i + D_i + 155\,\mathrm{dB}}{30\,\mathrm{dB} \cdot 160\,\mathrm{dB}} & \text{if } \zeta < \tilde{E}_i \leq D_i + 15\,\mathrm{dB} \\ -\frac{1}{160\,\mathrm{dB}} & \text{if } D_i + 15\,\mathrm{dB} < \tilde{E}_i \end{cases} \quad (22)$$

if $D_i + 15\,\mathrm{dB} < U_i + 10\,\mathrm{dB}$ (overlapping case).

These partial first-order derivatives are still not continuous at these points. This, however, does not influence the convergence itself but only its speed. Analogously, the partial second-order derivatives are defined as the left-sided derivative at their points of discontinuity.

3.5 Gain Computation

The time-varying gain factors $W_i'(k')$ are chosen such that

$$\tilde{E}_i(k') = \tilde{E}_i^{\mathrm{opt}}(k'), \quad (23)$$

which leads to the gain factor

$$W_i'(k') = 10^{[\tilde{E}_i^{\mathrm{opt}}(k') - E_i(k')]/20}. \quad (24)$$

In order to prevent hearing damage and pain, the gain is limited such that the resulting instantaneous equivalent

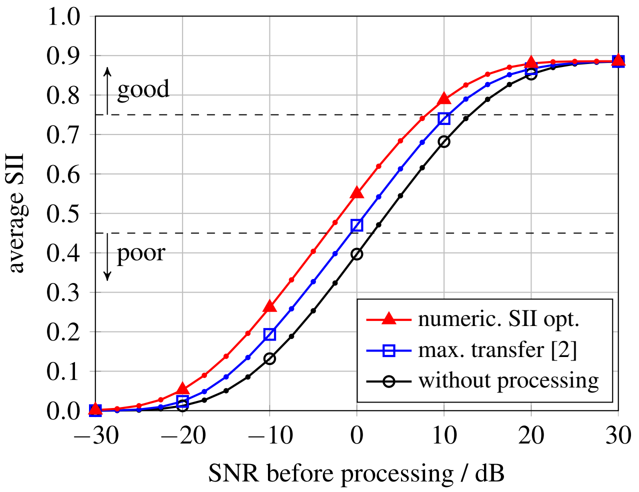


Figure 6: Comparison of average SII of speech at unchanged audio power disturbed by *factory1* noise.

spectrum level of the amplified speech in each subband does not exceed a maximum spectrum level $E_{\max} = 90\,\text{dB}$:

$$W_i(k') = \min\left\{ W_i'(k'), \sqrt{\frac{10^{E_{\max}/10}}{g_l^2 \cdot |S_i(k')|^2}} \right\}. \qquad (25)$$

The value of $E_{\max}$ is chosen in accordance to [11, Fig. 2.1].

4. RESULTS

The performance of the proposed algorithm was evaluated in terms of the SII using the so-called critical band procedure [7] for every speech file of the TIMIT database, in total 5.4 hours, disturbed by the *factory1* noise from the NOISEX-92 database at input signal-to-noise ratios (SNRs) between $-30\,\text{dB}$ and $30\,\text{dB}$ in steps of $2.5\,\text{dB}$ and at a sampling rate of $8\,\text{kHz}$.

Prior to processing, the speech database is scaled to match the overall sound pressure level of $62.35\,\text{dB}$ as specified in [7] for normal voice effort. The desired input SNR is achieved by adjusting the sound pressure level of the noise file in relation to $62.35\,\text{dB}$. 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 $20\,\text{ms}$ length. Finally, the average SII over all speech files is taken. Good communication systems have an SII of 0.75 or better while the SII of poor communication systems is below 0.45 [7].

In Figure 6 the average Speech Intelligibility Index is plotted after processing with

- the proposed numerically SII optimized algorithm,
- the method of maximal power transfer with adaptive noise floor as described in [2], and
- without processing.

It can be seen, that the proposed numerically SII optimized algorithm improves the average SII considerably while having the same average audio power. It also outperforms the method of maximal power transfer proposed in [2]. The subjective listening impression supports these results. Speech power is shifted towards higher frequencies, leading to a change of tone color but also subjectively improved intelligibility.

The numerical optimization takes about 3.5 iterations for low SNRs up to $5\,\text{dB}$ and about 14 iterations for high SNRs above $15\,\text{dB}$.

Sound samples and further information can be found at `http://www.ind.rwth-aachen.de/~bib/sauert10/`.

5. CONCLUSIONS

In this contribution, a new SII optimized near end listening enhancement algorithm is derived, with the constraint that the output signal has the same average audio power as the input signal. For this purpose, the calculation rules of the band audibility function are slightly modified to obtain a concave optimization function, which is then maximized using numerical optimization.

The instrumental evaluation by means of the SII has shown a noticeable better performance after processing with the proposed algorithm than without. Of course, considerable further improvements in SII can be achieved if additional audio power can be spent as shown, e. g., in [3].

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)* (Toulouse, France). May 2006, pp. 493–496. DOI: 10.1109/ICASSP.2006.1660065.

[2] Bastian Sauert, Gerald Enzner, and Peter Vary. "Near End Listening Enhancement with Strict Loudspeaker Output Power Constraining". In: *Proc. of Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)* (Paris, France). Sept. 2006.

[3] Bastian Sauert and Peter Vary. "Near End Listening Enhancement Optimized with Respect to Speech Intelligibility Index". In: *Proc. of European Signal Processing Conf. (EUSIPCO)* (Glasgow, Scotland). Vol. 17. Aug. 2009, pp. 1844–1848.

[4] Jong Won Shin et al. "Speech Reinforcement based on Partial Specific Loudness". In: *Proc. of European Conf. on Speech Communication and Technology (EUROSPEECH)* (Antwerp, Belgium). Aug. 2007, pp. 978–981.

[5] P. Shankar Chanda and Sungjin Park. "Speech Intelligibility Enhancement using Tunable Equalization Filter". In: *Proc. of Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)* (Honolulu, Hawaii, USA). Apr. 2007, pp. 613–616. DOI: 10.1109/ICASSP.2007.366987.

[6] Sungyub Yoo et al. "Speech Enhancement Based on Transient Speech Information". In: *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (New Paltz, NY, USA). Oct. 2005, pp. 62–65.

[7] American National Standards Institute. *Methods for the Calculation of the Speech Intelligibility Index*. ANSI S3.5-1997. 1997.

[8] 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). Vol. 8. Berlin [u.a.]: VDE-Verlag, Oct. 2008. ISBN: 978-3-8007-3120-6.

[9] Peter Vary. "An Adaptive Filterbank Equalizer for Speech Enhancement". In: *Signal Processing* 86.6 (June 2006): *Special Issue on Applied Speech and Audio Processing (dedicated to Prof. Hänsler)*, pp. 1206–1214. ISSN: 0165-1684. DOI: 10.1016/j.sigpro.2005.06.020.

[10] Heinrich W. Löllmann and Peter Vary. "Uniform and Warped Low Delay Filter-Banks for Speech Enhancement". In: *Speech Communication* 49 (July 2007): *Special Issue on Speech Enhancement*, pp. 574–587. ISSN: 0167-6393. DOI: 10.1016/j.specom.2007.04.009.

[11] Eberhard Zwicker and Hugo Fastl. *Psychoacoustics. Facts and Models*. 2nd ed. Berlin, Heidelberg, New York: Springer, 1999. ISBN: 978-3-540-65063-8.