Improving speech intelligibility in noise by SII-dependent preprocessing using frequency-dependent amplification and dynamic range compression

Henning Schepker[1], *Jan Rennies*[2], *Simon Doclo*[1,2]

[1]University of Oldenburg, Department of Medical Physics and Acoustics,
Signal Processing Group, Oldenburg, Germany
[2]Fraunhofer IDMT, Project Group Hearing-, Speech-, and Audio Technology,
Oldenburg, Germany
{henning.schepker,simon.doclo}@uni-oldenburg.de, jan.rennies@idmt.fraunhofer.de

Abstract

In this contribution, a new preprocessing algorithm to improve speech intelligibility in noise is proposed, which maintains the signal power before and after processing. The proposed *AdaptDRC* algorithm consists of two time- and frequency-dependent stages, which are both functions of the estimated SII. The first stage applies a time- and frequency-dependent amplification, while the second stage applies a time- and frequency-dependent dynamic range compression (DRC). Experiments with a competing speaker (CS) and a speech-shaped noise (SSN) show an increase in speech intelligibility for a wide range of SNRs for four different objective measures that are correlated with speech intelligibility. Listening tests conducted within the framework of the Hurricane Challenge with 175 subjects confirm these findings and show improvements of up to 20.5% in intelligibility for SSN and 12.3% for CS.

Index Terms: speech-in-noise enhancement, dynamic range compression, speech intelligibility

1. Introduction

Speech communication is gaining more and more importance in today's society. To ensure a high communication quality, a high speech intelligibility must be provided in speech communication applications. However, in many situations speech is degraded by additive noise and/or reverberation, leading to reduced intelligibility and increased listening effort [1, 2, 3]. Obviously, intelligibility can be maintained by raising the speech level to increase the signal-to-noise-ratio (SNR). This phenomenon can be observed in human speech production in noisy environments, which is called the Lombard effect [4]. For Lombard speech it has been found that not only the overall speech level is raised but also the spectral tilt is reduced [4], thereby increasing the energy at higher frequencies.

Although a simple amplification scheme can easily be implemented, it can lead to an overload of the amplification system or unpleasantly high sound levels. Consequently, it is desirable to increase speech intelligibility while maintaining equal powers of both the unprocessed and the processed speech signal.

One of the first attempts to investigate the influence of different signal preprocessing strategies on speech intelligibility was done by Licklider and Pollack [5]. Although they only investigated speech in quiet, they reported no degradation of speech intelligibility for various types of filtering and clipping strategies as well as their combinations.

A more recent approach based only on spectral shaping was proposed by Sauert and Vary, where a time- and frequency-dependent amplification of the speech signal was carried out aiming to maximize the Speech Intelligibility Index (SII) [6, 7]. This approach however suffers from spectral adaptation to the noise characteristics [8], leading them to use a transition between an SII-weighting and unity-weighting.

As another approach Niederjohn and Grotelueschen [9] proposed to use high-pass filtering followed by a static rapid amplitude compression. They reported an enhanced speech intelligibility in white noise over the unprocessed signal at the same SNR. Zorila et al. [10] recently adopted part of this approach in their *SSDRC* algorithm, in which they combined a speech-signal-dependent spectral shaping and a static broadband dynamic range compression (DRC) scheme. They reported improvements in SSN and CS for three different SNRs.

The abovementioned approaches use either only frequency shaping or additional static compression characteristics. By applying static compression characteristics, the speech signal is also changed in case of good intelligibility. On the contrary, the *AdaptDRC* algorithm proposed in this paper combines these two stages, where both the amplification stage as well as the compression characteristics are based on the estimated SII, leading to no changes in case of good intelligibility.

This paper is organized as follows. In Section 2 the considered scenario and some definitions will be given. Section 3 provides a detailed overview of the proposed *AdaptDRC* algorithm. Results from evaluations using objective measures as well as from a formal listening test are given in Section 4.

2. Scenario and definitions

Consider the acoustic scenario depicted in Figure 1. The unprocessed speech signal $s[k]$ at discrete time k is modified using the weighting function $W\{\cdot\}$ and played back via a loudspeaker. A microphone picks up the disturbed speech signal $y[k]$, which is the mixture of the modified speech signal $\tilde{s}[k]$ convolved with the room impulse response $h[k]$ and the additive noise disturbance $r[k]$, i.e.

$$y[k] = \tilde{s}[k] * h[k] + r[k], \tag{1}$$

where $*$ denotes convolution. An estimate $\hat{r}[k]$ of the noise signal $r[k]$ can be obtained by using e.g. adaptive filtering techniques to model the room impulse response $h[k]$ [11]. Using the estimated noise signal $\hat{r}[k]$, the estimated room impulse response $\hat{h}[k]$, and the clean speech signal $s[k]$, the processed

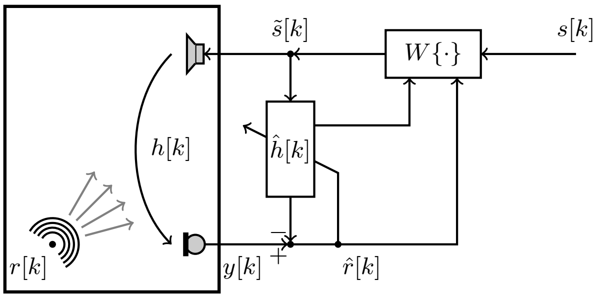

Figure 1: Considered acoustical scenario.

speech signal $\tilde{s}[k]$ is then computed as

$$\tilde{s}[k] = W\{s[k],\hat{r}[k],\hat{h}[k]\}s[k]. \tag{2}$$

In the following we assume that a perfect noise estimate is available, i.e. $\hat{r}[k] = r[k]$, and no reverberation is present, i.e. $h[k] = \delta[k]$. We aim at finding a weighting function $W\{\cdot\}$ that enhances the intelligibility of $\tilde{s}[k]+r[k]$ compared to $s[k]+r[k]$ under an equal power constraint.

The signal $s[k]$ is first split into N subband signals $s_n[k]$, $n = 1,\ldots,N$, using a real-valued filterbank. In our implementation, we have used an all-pass filterbank based on doubly-complementary IIR filters [12], splitting the signals into octave bands with center frequencies ranging from 125 Hz to 16 kHz ($N = 8$). Additionally, each subband signal $s_n[k]$ is framed into non-overlapping blocks of length M, i.e. $s_n^l[m] = s_n[lM+m]$, $m = 0,\ldots,M-1$ with block index l. The power of the lth block in the nth subband is equal to

$$\phi_n[l] = \frac{1}{M}\sum_{m=0}^{M-1}(s_n^l[m])^2. \tag{3}$$

The broadband power $\phi[l]$ in the lth block is defined as the sum over all subbands n. Furthermore, the equivalent speech spectrum level is defined as [13]

$$e_n[l] = 10\log_{10}\left(\frac{\phi_n[l]}{\Delta f_n}\right), \tag{4}$$

with Δf_n the bandwidth of the nth subband. Similarly, the equivalent disturbance spectrum level is defined as

$$d_n[l] = 10\log_{10}\left(\frac{1}{M\Delta f_n}\sum_{m=0}^{M-1}(\hat{r}_n^l[m])^2\right), \tag{5}$$

where $\hat{r}_n^l[m]$ is defined similarly as $s_n^l[m]$.

3. Algorithm

The proposed *AdaptDRC* algorithm is schematically depicted in Figure 2. It combines a time- and frequency-dependent amplification stage with a time- and frequency-dependent DRC stage, where both the amplification and the DRC are functions of the estimated SII. The SII estimation is adopted from [7] but slightly changed here. The SII in the lth block is estimated from the clean speech signal $s[k]$ and the noise signal $\hat{r}[k]$ as

$$\hat{SII}[l] = \sum_{n=1}^{N} i_n \hat{a}\left(e_n[l],d_n[l]\right), \tag{6}$$

where i_n denotes the weighting of the nth octave subband according to [13] and $\hat{a}\left(e_n[l],d_n[l]\right)$ approximates the audibility

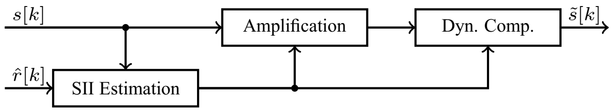

Figure 2: Schematic flow-graph of the *AdaptDRC* algorithm

function $a\left(e_n[l],d_n[l]\right)$ defined in [13], which estimates the audibility of the speech signal relative to the noise signal as

$$\hat{a}\left(e_n[l],d_n[l]\right) = q\left(e_n[l],d_n[l]\right)\hat{g}\left(d_n[l]\right). \tag{7}$$

The function $q\left(e_n[l],d_n[l]\right)$ is a mapping function of the SNR according to

$$q\left(e_n[l],d_n[l]\right) = \max\left\{\min\left\{\frac{e_n[l]-d_n[l]+15\,dB}{30\,dB},1\right\},0\right\} \tag{8}$$

and the function $\hat{g}\left(d_n[l]\right)$ describes the distortion of speech with louder levels assuming the special case of $e_n[l] = d_n[l] + 15\,\mathrm{dB}$ as

$$\hat{g}\left(d_n[l]\right) = \min\left\{1 - \frac{d_n[l]+15\,dB-u_n-10\,dB}{160\,dB},1\right\}, \tag{9}$$

which depends on the so-called standardized equivalent speech spectrum level u_n defined in [13]. Note that $\hat{a}\left(e_n[l],d_n[l]\right)$ differs from the approximation of the audibility function proposed in [7] in that it takes into account the flooring and ceiling effect of the function $q\left(e_n[l],d_n[l]\right)$, which was neglected in [7]. Additionally, by using the approximation $\hat{g}\left(d_n[l]\right)$ it is assumed that $e_n[l] = d_n[l] + 15\,\mathrm{dB} < u_n + 170\,\mathrm{dB}$, which is however valid in nearly all conditions.

3.1. Amplification

The amplification stage of the *AdaptDRC* algorithm aims at increasing the power in the high-frequency regions of the speech signal. This is achieved by adopting the approach proposed in [8]. In [8] an amplification of the subband speech signal according to their contribution to speech intelligibility as described within the SII is used. In contrast, we assume that most noise signals $r[k]$ have a spectral slope that is steeper than or equal to the spectral slope commonly observed in speech signals $s[k]$. Therefore, in the amplification stage of the *AdaptDRC* algorithm we focus on amplifying high-frequency regions by uniformly distributing the speech signal power in all subbands in case of low speech intelligibility as predicted by (6), leading to a reduced spectral slope of $\tilde{s}[k]$ compared to $s[k]$. The amplification gain of the lth block is computed as

$$w_n[l] = \sqrt{\frac{\phi_n^{\hat{SII}[l]}[l]}{\sum_{\lambda=1}^{N}\phi_\lambda^{\hat{SII}[l]}[l]}\cdot\frac{\phi[l]}{\phi_n[l]}}. \tag{10}$$

If the estimated SII is equal to one, equation (10) results in unity gain, while for $\hat{SII}[l] = 0$ a uniform distribution for the output powers of all subbands in the lth block is achieved.

3.2. Dynamic range compression

The DRC stage of the *AdaptDRC* algorithm aims at amplifying low-levels signals that are assumed to be not well audible and attenuating high-level signal that are assumed to be well audible. Usually, in DRC fixed compression ratios are applied to the broadband or the subband signals. However, in case $s_n[k]$ is already well audible, DRC can lead to signal degradation that may influence the perceived quality or speech intelligibility. Therefore, for the *AdaptDRC* algorithm we propose to change the compression characteristic over time, depending on the speech and noise levels.

To compute the time-varying compressive gain, first an estimate of the envelope of each subband signal is computed, i.e.

$$\bar{s}_n[k] = \begin{cases} \alpha_a \bar{s}_n[k-1] + (1-\alpha_a)\,|s_n[k]| \\ \qquad \text{if} \quad |s_n[k]| \geq \bar{s}_n[k-1] \\ \alpha_r \bar{s}_n[k-1] + (1-\alpha_r)\,|s_n[k]| \\ \qquad \text{if} \quad |s_n[k]| < \bar{s}_n[k-1] \end{cases}, \qquad (11)$$

where α_a and α_r are attack and release smoothing constants. In DRC algorithms the processed speech signal in the lth block is computed as

$$\tilde{s}_n^l[m] = s_n^l[m] p\left(\boldsymbol{\lambda}_n[l], (\bar{s}_n^l[m])^2\right), m = 0, \ldots, M-1, \qquad (12)$$

where $p\left(\boldsymbol{\lambda}_n[l], (\bar{s}_n^l[m])^2\right)$ in the dB-domain linearly inter- and extrapolates the input-output-characteristic (IOC), which is defined by the parameter vector $\boldsymbol{\lambda}_n[l]$ containing a set of input and output powers $\gamma_{n,i}[l]$ and $\xi_{n,i}[l]$. In our algorithm we have used three input and output powers to define the IOC, i.e.

$$\boldsymbol{\lambda}_n[l] = \left[\gamma_{n,1}[l]\ \gamma_{n,2}[l]\ \gamma_{n,3}[l]\ \xi_{n,1}[l]\ \xi_{n,2}[l]\ \xi_{n,3}[l]\right]. \quad (13)$$

The novelty of the proposed *AdaptDRC* algorithm in comparison to previously proposed algorithms [9, 10] is the fact that the IOC changes for each block l, depending on the speech and noise characteristics. The proposed IOC of the *AdaptDRC* algorithm is defined as

$$\begin{aligned} \gamma_{n,1}[l] &= 1 \\ \gamma_{n,2}[l] &= \phi_n[l] \\ \gamma_{n,3}[l] &= \nu \end{aligned} \qquad (14)$$

and

$$\begin{aligned} \xi_{n,1}[l] &= \phi_n^{1-cr_n[l]}[l] \\ \xi_{n,2}[l] &= \phi_n[l] \\ \xi_{n,3}[l] &= \phi_n^{1-cr_n[l]}[l]\nu \end{aligned} \qquad (15)$$

where ν is a conversion constant from dB FS to dB SPL and $cr_n[l]$ is the time-varying compression ratio, defined as

$$cr_n[l] = \max\left\{cr_{(\max)} \cdot (1 - q\,(e_n[l], d_n[l])), 1\right\}, \qquad (16)$$

where $cr_{(\max)}$ defines the maximum compression ratio. For subband SNRs lower than or equal to -15 dB $q(e_n[l],d_n[l]) = 0$, such that $cr_{(\max)}$ will be applied, while for subband SNRs larger than or equal to +15 dB $q(e_n[l],d_n[l]) = 1$, such that no compression will be applied. In Figure 3 exemplary IOCs for the proposed *AdaptDRC* algorithm are shown, assuming that $\phi_n[l] = 10^{50/10}$. For case 1 we assume a sufficiently high SNR resulting in $cr_n[l] = 1$ and hence no compression is applied. For case 2 $cr_n[l] = cr_{(\max)} = 8$ is assumed, i.e. a sufficiently low SNR is present, which leads to large amplifications for low input levels and strong attenuations of high input levels.

Furthermore, the amplification stage discussed in Section 3.1 can be directly incorporated into the IOC by redefining $\xi_{n,i}[l]$ in (15) as

$$\begin{aligned} \xi_{n,1}[l] &= \phi_n^{1-cr_n[l]}[l] w_n[l] \\ \xi_{n,2}[l] &= \phi_n[l] w_n[l] \\ \xi_{n,3}[l] &= \phi_n^{1-cr_n[l]}[l] w_n[l] \nu. \end{aligned} \qquad (17)$$

This is depicted in Figure 3 as case 3, assuming that $w_n = 10^{15/20}$ and $cr_n[l] = 8$. The additional amplification results in a shift of the IOC towards higher output levels.

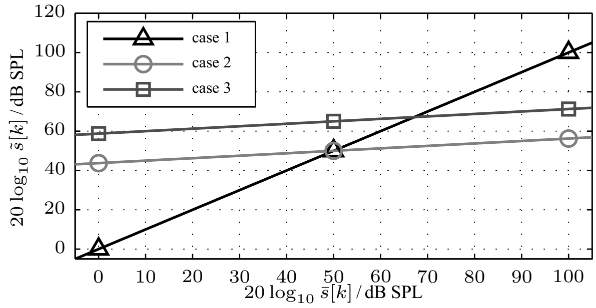

Figure 3: Exemplary IOC for three different cases. Symbols indicate the points defined by the input and output powers $\gamma_{n,i}[l]$ and $\xi_{n,i}[l]$ and lines indicate $p(\boldsymbol{\lambda}_n[l], \bar{s}_n^l[m])$. The following cases are shown: case 1: $cr_n[l] = 1$; case 2: $cr_n[l] = 8$; case 3: $w_n = 10^{15/20}$ and $cr_n[l] = 8$.

3.3. Smoothing of gain functions

Directly using the gain derived from the IOC defined in equations (14) and (17) may lead to noticeable processing artifacts. To counteract these artifacts a two-stage smoothing procedure is applied. First, the IOC is smoothed, i.e.

$$\bar{\lambda}_{n,j}[l] = \alpha_b \bar{\lambda}_{n,j}[l-1] + (1-\alpha_b)\lambda_{n,j}[l], \qquad (18)$$

where $\lambda_{n,j}[l]$ is the jth element of the IOC parameter vector $\boldsymbol{\lambda}_n[l]$ and α_b is a smoothing constant. Second, the resulting gain is recursively smoothed with smoothing factor α_p, i.e.

$$\begin{aligned} \bar{p}\left(\bar{\boldsymbol{\lambda}}_n[l], (\bar{s}_n^l[m])^2\right) &= \alpha_p \bar{p}\left(\bar{\boldsymbol{\lambda}}_n[l], (\bar{s}_n^l[m-1])^2\right) \\ &\quad + (1-\alpha_p)\, p\left(\bar{\boldsymbol{\lambda}}_n[l], (\bar{s}_n^l[m])^2\right). \end{aligned} \qquad (19)$$

The processed subband signals $\tilde{s}_n^l[m]$ are then obtained by applying (19) instead of $p\left(\boldsymbol{\lambda}_n[l], (\bar{s}_n^l[m])^2\right)$ in (12). The application of (19) however typically leads to changes in the broadband signal power and therefore does not satisfy the equal power constraint. While in an offline procedure the signal could be easily rescaled after processing to satisfy the constraint, this is not possible in an online application. Therefore, after applying the inverse filterbank, an additional broadband gain is applied that yields approximately equal powers. This gain is calculated by dividing the smoothed versions of the broadband input and output powers with smoothing constant α_L.

4. Results

The proposed *AdaptDRC* algorithm has been evaluated using several objective measures and using a formal subjective listening test as part of the Hurricane Challenge [14]. Speech material was taken from the Harvard speech corpus recorded by one male native British English speaker [15]. Two different noise types were used in the evaluation at three different SNRs each, namely:

- Competing Speaker (CS) at SNRs: $-21, -14, -7$ dB
- Speech Shaped Noise (SSN) at SNRs: $-9, -4, +1$ dB.

These SNRs were selected to yield keyword correct scores of approximately 25%, 50% and 75% in the unprocessed *Reference* condition . The speech and noise signals were sampled at $f_s = 16$ kHz. In the results presented hereafter, the parameters according to Table 1 were used, which yielded a good compromise in terms of intelligibility improvement for both noises based on informal listening tests. Note that for clarity

Table 1: Integration constants of the different smoothing parameters and the block length used in the evaluation.

Parameter	Value	Parameter	Value
τ_a	0.025 s	τ_w	0.250 s
τ_r	0.500 s	τ_b	0.250 s
τ_L	0.250 s	M	0.020 s
$cr_{(\max)}$	8		

the corresponding integration time constants are given, while in the description of the algorithm the smoothing constants were used. Sound examples can be found at http://www.sigproc.uni-oldenburg.de/audio/adaptdrc/is2013.html.

4.1. Objective measures

To quantify the effect of the proposed *AdaptDRC* algorithm on speech intelligibility, we have used four objective measures that have shown high correlation with speech intelligibility in previous experiments, namely the Speech Transmission Index (STI) [16], the SII [13], the Extended SII (ESII) [17] and the Short Time Objective Intelligibility Measure (STOI) [18]. For all 240 sentences of the Harvard speech corpus the objective measures have been calculated in all SNR and noise conditions. Note that although in [17] it is originally proposed to use SSN as speech input to the ESII, we have used the speech utterances as speech input, thus differences in ESII and SII can be expected also for the SSN. Figure 4 shows the results of the objective evaluation. For all considered objective measures, an improvement of the processed speech over the unprocessed speech can be observed, thereby indicating the effectivity of the *AdaptDRC* algorithm. Comparing both noises, STI, SII and STOI predict a larger intelligibility improvement for the SSN than for CS, while ESII predicts a smaller intelligibility improvement, which appears counterintuitive. Furthermore, the predicted absolute intelligibility improvement is largest for the STI and SII, which consider only long-term spectral information of the speech and noise signals. For ESII and STOI, which both consider short-term spectral information of the speech and noise signals, smaller absolute improvements are predicted.

4.2. Listening test

A subjective listening test was conducted as part of the 2013 Hurricane Challenge [14] with 175 native and audiologically normal-hearing subjects. In contrast to the objective evaluation, a subset containing only the first 180 sentences of the Harvard corpus was used. Figure 5 shows the results for the *AdaptDRC* algorithm and the unprocessed *Reference* in terms of percentage correctly understood keywords as a function of SNR. The results show that the proposed *AdaptDRC* algorithm increases the speech intelligibility for both noise conditions and all SNRs. Increases of up to approximately 12% and 20% can be achieved for CS and SSN, respectively (see Table 2). As expected, a larger increase can be achieved for the stationary SSN compared to the instationary CS. A statistical analysis of the data was carried out to confirm these findings. Since Shapiro-Wilk tests revealed non-normality of the data, an aligned rank transform procedure [19] was carried out to replace a conventional 3-factor (processing, SNR, noise) repeated-measures ANOVA. Note that, although the SNRs in both noises were not physically the same, they were regarded as high, mid and low in the analysis. The 3-factor analysis revealed significant effects ($p < 0.01$) of all factors and significant interaction ($p < 0.05$) between all factor combinations except processing×noise, confirming the

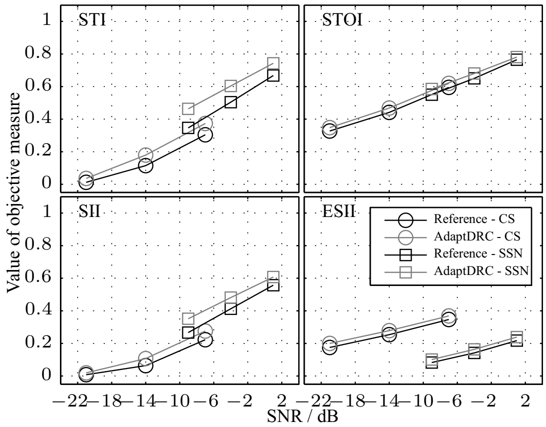

Figure 4: Objective measures for both noise types (CS and SSN).

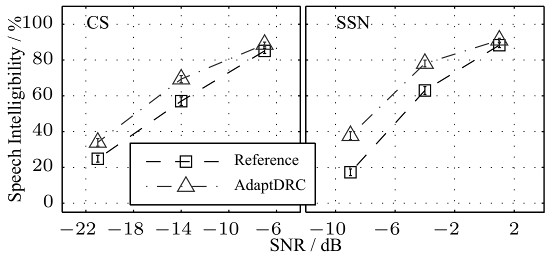

Figure 5: Results from listening test for both noise types (CS and SSN). Error-bars indicate standard errors of the mean values.

Table 2: Results from formal listening test.

Noise	SNR	Processed	Reference	Gain
CS	-7 dB	88.7 ± 1.3%	85.1 ± 1.5%	3.6%
	-14 dB	69.3 ± 2.0%	57.0 ± 2.4%	12.3%
	-21 dB	34.7 ± 2.1%	24.8 ± 1.9%	9.2%
SSN	+1 dB	91.1 ± 1.3%	88.3 ± 1.3%	2.8%
	-4 dB	78.1 ± 1.8%	63.0 ± 2.2%	15.1%
	-9 dB	37.7 ± 2.2%	17.3 ± 1.8%	20.5%

previous findings.

5. Conclusions

In this paper, a novel SII-dependent preprocessing strategy was proposed that combines a time- and frequency-dependent amplification stage and a time- and frequency-dependent DRC stage to improve speech intelligibility in noise. Evaluations using four different objective measures showed improvements over the unprocessed *Reference* for both CS and SSN. Results from subjective listening tests confirm these findings and show significant improvements of up to 20% in speech intelligibility in the entire range of SNRs. The results show that the *AdaptDRC* algorithm is capable of enhancing speech intelligibility both for stationary as well as for time-varying maskers with speech-like spectra.

6. Acknowledgements

The authors thank Francois X. Nsabimana for helping to prepare the sound files and gratefully acknowledge the efforts of the organizers of the 2013 Hurricane Challenge. This work was supported in part by the Cluster of Excellence 1077 "Hearing4All", funded by the German Research Foundation (DFG).

7. References

[1] Morimoto, M., Sato, H., and Kobayashi, M., “Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces”, J. Acoust. Soc. Am. 116(3):1607-1613, 2004.

[2] George, E. L. J., Goverts, S. T., Festen, J.M., and Houtgast, T., “Measuring the effects of reverberation and noise on sentence intelligibility for hearing-impaired listeners”, J. Speech Lang. Hear. Res. 53:1429-1439, 2010.

[3] Schepker, H., Rennies, J., Holube, I. and Kollmeier, B., “Relationship of listening effort, speech intelligibility and STI” (in German “Zusammenhang von Höranstrengung, Sprachverständlichkeit und STI”), in: Fortschritte der Akustik - DAGA 2012, Darmstadt, Germany, Mar. 2012, pp. 875-876.

[4] Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I. and Stokes, M. A., “Effects of noise on speech production: Acoustic and perceptual analyses”, J. Acoust. Soc. Am., 84(3):917-928, 1988.

[5] Licklider, J. C. R. and Pollack, I., “Effects of Differentiation, Integration, and Infinite Peak Clipping upon the Intelligibility of Speech”, J. Acoust. Soc. Am., 20(1):42-51, 1948.

[6] Sauert, B. and Vary, P., “Near end listening enhancement optimized with respect to speech intelligibility index and audio power limitations”, in: Proc. of European Signal Processing Conference (EUSIPCO). Aalborg, Denmark, 2010, pp. 1919-1923.

[7] Sauert, B. and Vary, P., “Recursive Closed-Form Optimization of Spectral Audio Power Allocation for Near End Listening Enhancement”, in: Proc. of ITG-Fachtagung Sprachkommunikation, Bochum, Germany, 2010.

[8] Sauert, B. and Vary, P., “Near-end listening enhancement in the presence of bandpass noises”, in: Proc. of ITG-Fachtagung Sprachkommunikation, Braunschweig, Germany, 2012, pp. 195-198.

[9] Niederjohn, R. and Grotelueschen, J., “The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression”. IEEE Trans. Acoust. Speech., 24(4):277-282, 1976.

[10] Zorila, T.-C., Kandia, V. and Stylianou, Y., “Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression”, in: Proc. of Interspeech, Portland, USA, 2012.

[11] Benesty, J., “Springer handbook of speech processing”, Springer Verlag, 2008.

[12] Regalia, P., Mitra, S., Vaidyanathan, P., Renfors, M. and Neuvo, Y., “Tree-structured complementary filter banks using all-pass sections”, IEEE Trans. Circuits Syst., 34(12):1470-1484, 1987.

[13] ANSI, “Methods for Calculation of the speech intelligibility index”, American National Standard ANSI S3.5-1997, New York, USA, 1997.

[14] Cooke, M., Mayo, C., and Valentini-Botinhao C., ”Intelligibility-enhancing speech modifications: the Hurricane Challenge“, submitted to Interspeech, Lyon, France, 2013.

[15] Rothauser, E.H., Chapman, W.D., Guttman, N., Silbiger, H.R., Hecker, M.H.L., Urbanek, G.E., Nordby, K.S., Weinstock, M., “IEEE Recommended practice for speech quality measurements”, IEEE Trans. Acoust Speech, 17:225-246, 1969.

[16] IEC, “Sound System Equipment - Part 16: Objective rating of speech intelligibility by speech transmission index”, International Standard IEC 60268-16 (International Electrotechnical Commission), Geneva, Switzerland, 2003.

[17] Rhebergen, K. S. and Versfeld, N. J., “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(4):2181-2192, 2005.

[18] Taal, C., Hendriks, R., Heusdens, R. and Jensen, J., “An algorithm for intelligibility prediction of time frequency weighted noisy speech”, IEEE Trans. Audio Speech, 19(7):2125-2136, 2011.

[19] Wobbrock, J. O., Findlater, L., Gergle, D. ad Higgins, J. J., “The aligned rank transform for nonparametric factorial analyses using only anova procedures”. In: Proc. of SIGCHI Conference on Human Factors in Computing Systems, Vancouver, Canada, 2011, pp. 143-146