

# Adaptive compressive onset-enhancement for improved speech intelligibility in noise and reverberation

Felicitas Bederna<sup>1</sup>, Henning Schepker<sup>2</sup>, Christian Rollwage<sup>1</sup>, Simon Doclo<sup>1,2</sup>, Arne Pusch<sup>1</sup>, Jorg Bitzer<sup>1,3</sup>, Jan RENNIES<sup>1</sup>

<sup>1</sup>Fraunhofer Institute for Digital Media Technology IDMT, Division Hearing, Speech and Audio Technology, and Cluster of Excellence Hearing4all, Oldenburg, Germany

<sup>2</sup>Signal Processing Group, Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4all, University of Oldenburg, Germany

<sup>3</sup>Institute of Hearing Technology and Audiology (IHA), Jade-University of Applied Sciences Wilhelmshaven / Oldenburg / Elsfleth, Germany

felicitas.bederna@idmt.fhg.de, jan.rennies@idmt.fhg.de

## Abstract

Near-end listening enhancement (NELE) algorithms aim to pre-process speech prior to playback via loudspeakers so as to maintain high speech intelligibility even when listening conditions are not optimal, e.g., due to noise or reverberation. Often NELE algorithms are designed for scenarios considering either only the detrimental effect of noise or only reverberation, but not both disturbances. In many typical applications scenarios, however, both factors are present. In this paper, we evaluate a new combination of a noise-dependent and a reverberation-dependent algorithm implemented in a common framework. Specifically, we use instrumental measures as well as subjective ratings of listening effort for acoustic scenarios with different reverberation times and realistic signal-to-noise ratios. The results show that the noise-dependent algorithm also performs well in reverberation, and that the combination of both algorithms can yield slightly better performance than the individual algorithms alone. This benefit appears to depend strongly on the specific acoustic condition, indicating that further work is required to optimize the adaptive algorithm behavior.

**Index Terms:** speech intelligibility, near-end listening enhancement, reverberation

## 1. Introduction

In many applications speech played back via loudspeakers is degraded by noise and reverberation in the listening room resulting in reduced speech understanding [1, 2]. In order to maintain high speech intelligibility even in adverse conditions, it is possible to pre-process the loudspeaker signal using so called near-end listening enhancement (NELE) algorithms. A simple approach would be to increase the speech level, when noise and reverberation are present. However, while this may increase speech intelligibility, it may also overload the amplification system or can become unpleasantly loud. Hence, it is desirable to design algorithms that maintain the same signal power before and after processing.

Many approaches have been presented to improve speech understanding under an equal-power constraint [3, 4, 5, 6]. Most of these approaches, however, only take into account the presence of either noise [7, 8, 9] or reverberation [10, 11, 12, 13, 14] in their design. While this typically eases the design, in practice both disturbances are present at the same time. One solution could be to design NELE algorithms taking into ac-

count the combined effect of both noise and reverberation (e.g. [15, 16]). These approaches incorporated the detrimental effect of reverberation by adding the late reverberation [16] or both late reverberation and early reflections [15] as an additional noise term in a model-based approach. On the other hand, it is also possible that algorithms designed for either noise or reverberation may also provide enough improvement in scenarios that are both noisy and reverberant, hence making a combined approach obsolete.

Previously, the benefit of NELE algorithms was commonly evaluated using speech intelligibility tests in noise-only scenarios [4]. However, this usually requires very low signal-to-noise ratios (SNRs) to show differences between algorithms, since even unprocessed speech is often perfectly understood at, e.g., SNRs above -3 dB [2, 17]. In contrast, realistic scenarios often exhibit much more favorable SNRs, e.g., generally positive SNRs where speech intelligibility is already at ceiling [18]. In [19] it was shown that listening effort can be used to assess the benefit of NELE algorithms at much higher SNRs, which are more representative for everyday-scenarios.

In this paper we evaluate two different NELE algorithms in noisy and reverberant environments and propose a combination of both algorithms in a common framework. More specifically, we use the noise-dependent AdaptDRC algorithm [7] and the reverberation-dependent OE algorithm [14]. In contrast to previous studies using speech intelligibility measurements at rather low SNRs, we also consider more realistic SNRs and different reverberation times and use subjectively rated listening effort [20] as well as instrumental measures to quantify the benefit of both algorithms and their combination.

## 2. Near-end listening enhancement

The general NELE problem is to modify the unprocessed (clean) speech signal  $s[k]$  at discrete time  $k$  to produce the modified speech signal  $\tilde{s}[k]$  before playing it back via a loudspeaker. In the listening room, additive noise  $r[k]$  and reverberation characterized by a room impulse response (RIR)  $h[k]$  disturb the speech information transmission. The goal of a NELE algorithm is to process the speech such that  $\tilde{s}[k] * h[k] + r[k]$  is more intelligible than  $s[k] * h[k] + r[k]$ , where  $*$  denotes convolution. In practice an estimate  $\hat{r}[k]$  of the noise signal  $r[k]$  as well as an estimate  $\hat{h}[k]$  of the RIR  $h[k]$  can be obtained by using, e.g., adaptive filtering techniques to model  $h[k]$  [21].

In the following, we assume perfect knowledge of the RIR (i.e.,  $\hat{h}[k] = h[k]$ ) and hence also a perfect noise estimate to be available (i.e.,  $\hat{r}[k] = r[k]$ ).

In this paper we consider the noise-dependent AdaptDRC algorithm [7] and the reverberation-dependent OE algorithm [14]. Since these algorithms focus only on the effect of either noise or reverberation, we propose to combine them in a joint framework.

## 2.1. AdaptDRC

The AdaptDRC algorithm [7] was designed to improve speech intelligibility in noisy environments by combining a frequency-shaping stage aiming to enhance high frequency components and a dynamic range compression (DRC) stage to amplify low-level speech components while reducing high-level speech components. A brief description of the algorithm is provided below (for more details, see [7]).

The algorithm processes speech signals  $s(k)$  in time frames of 20 ms. In the following,  $l$  and  $m$  indicate the frame index and the discrete time index within a frame, respectively. The speech frame  $s^l(m)$  and the noise frame  $r^l(m)$  are divided into  $N = 8$  octave bands centered at 125 Hz to 16 kHz. In the frequency-shaping stage, a time- and frequency-dependent amplification  $w_n(l)$  is applied, with  $n$  denoting the band index. Using these octave-band signals a simplified version of the Speech Intelligibility Index (SII) is estimated for each frame  $l$ . Based on the estimated SII  $\hat{SII}(l)$ , a weighting function is applied that results in an adaptive amplification behavior. In case of a low predicted speech intelligibility, i.e.  $\hat{SII}(l) \rightarrow 0$ , a uniform distribution of the speech power across all subbands is applied, which leads to an increase of speech power in high frequency regions. For high SII, i.e.,  $\hat{SII}(l) \rightarrow 1$ , no processing is applied, i.e.,  $w_n(l) = 1$ . In the DRC stage the goal is to amplify low-level parts of each speech subband relative to high-level parts in order to increase their audibility. To avoid signal distortions for high SNRs, an adaptive compression scheme is used, where for each frame a subband-dependent compression ratio is computed based on the SNRs. For subband SNRs  $\geq 15$  dB a compression ratio of 1 : 1 is applied, while for subband SNRs  $\leq -15$  dB a maximum compression ratio of 8 : 1 is applied. For intermediate subband-SNRs a continuous linear transition of the compression ratio is used. After recombination of the subband signals, the rms-power of the re-combined output signal is adaptively normalized to the (broadband) input rms-power to meet the equal-power constraint. Note that in very favorable listening conditions the two stages of the algorithm do not modify the input signal, while in very unfavorable conditions the maximum degree of processing is applied. An example is shown in the top right panel of Figure 2.

## 2.2. Onset-Enhancement and Overlap-Masking Reduction (OE)

The OE algorithm [14] was designed to increase speech intelligibility in reverberant environments by reducing the influence of self-masking of speech by its own reflections. A brief description of the algorithm is provided below (for more details, see [14], where it is described as algorithm 2).

First, the speech  $s(k)$  is split into time frames of 32 ms and transformed into the frequency domain using the fast Fourier transform (FFT), yielding  $S(f, l)$  for the  $f$ -th frequency bin and the  $l$ -th block. To reduce the amount of self-masking due to the smearing of reverberated speech, the goal is to increase the

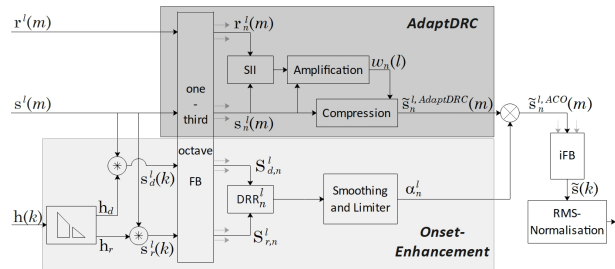


Figure 1: Block diagram of the ACO algorithm.

consonant-vowel-ratio. This increase is controlled by a time- and frequency-dependent direct-to-reverberant ratio of running speech (DRRs). To compute the DRRs, the RIR is assumed to be a linear and time-invariant system  $h(k) = h$  and is separated into a direct component  $h_d$  and a reverberant component  $h_r$  (i.e.,  $h = h_d + h_r$ ) using a separation constant of 1.3 ms, and transformed to the frequency domain, yielding  $H(f) = H_d(f) + H_r(f)$ .

Using this separation, the direct sound and the reverberant component of the speech in the frequency-domain can be computed as  $S_d(f, l) = S(f, l)H_d(f)$  and  $S_r(f, l) = S(f, l)H_r(f)$ . In order to take into account the spectral resolution of the human auditory system, the speech spectra  $S_d(f, l)$  and  $S_r(f, l)$  are grouped into subbands of an equivalent rectangular bandwidth approximating that of auditory filters. Using this grouping, the energy of the direct speech component  $\phi_{s_d}(n, l)$  and the energy of the reverberant speech component  $\phi_{s_r}(n, l)$  in the  $n$ th auditory filter are computed, their ratio representing the block- and filter-dependent  $DRR(n, l)$ .

These DRRs are utilized to enhance the consonant-to-vowel ratio based on the observation that the DRR is typically larger for consonants than for vowels. A frequency-dependent gain  $\alpha_n^l$  is computed, where frames with higher DRRs are amplified, while frames with lower DRRs are attenuated. To maintain the broadband power of the input signal, an additional broadband normalization is applied. In the presence of no reverberation, the applied gain is perfectly compensated for by the renormalization, and hence no processing is applied. An example is shown in the bottom left panel of Figure 2.

## 2.3. Adaptive compressive onset enhancement (ACO)

The goal of the proposed ACO algorithm is to improve speech intelligibility in both noisy and reverberant environments. It combines the concepts of the noise-dependent AdaptDRC algorithm and the reverberant-dependent OE algorithm as depicted in Figure 1. While the AdaptDRC algorithm and OE algorithm outlined in Section 2.1 and Section 2.2 use different filterbank structures (octave-band vs FFT) and frame lengths (20 ms vs 32 ms), the ACO algorithm processes a speech signal in 50% overlapping time frames of 32 ms and uses a common 21-band one-third octave filter bank with center frequencies from 160 Hz to 16 kHz. The amplification factors  $w_n(l)$  of the AdaptDRC algorithm and  $\alpha_n^l$  of the OE algorithm are calculated separately for each speech frame. The amplification and compression stages of the AdaptDRC algorithm are applied to the speech frame first, before multiplying the processed frequency bands with the corresponding gain factors of the OE algorithm. After applying the amplification gains, the signal  $\hat{s}_n^{l, ACO}(m)$  is reconstructed to the processed speech signal  $\hat{s}(k)$  with an inverse

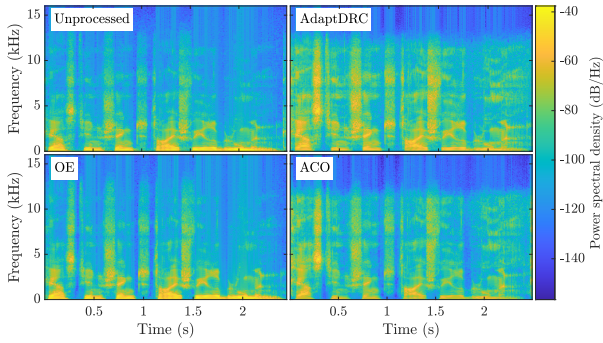


Figure 2: Exemplary spectrogram of unprocessed speech (top left) and speech processed by the three algorithms (calculated with an RIR with  $T_{30} = 4, 14$  s and noise at  $-20$  dB SNR).

filterbank, followed by a normalization to obtain the broadband signal power of the original speech.

An illustration of the ACO processing is provided in Figure 2 (bottom right panel). The high-frequency amplification from the AdaptDRC-component can clearly be seen in comparison to unprocessed speech (top left panel), although it is less pronounced for ACO-processed speech than for AdaptDRC-processed speech due to the interaction with the OE-processing. The impact of the OE-component is especially visible as sharper temporal modulations (or more clearly pronounced speech pauses, e.g., at  $t \approx 1$  s, or 1.8 s).

### 3. Evaluation

#### 3.1. Methods

Evaluations were conducted based on two instrumental measures as well as subjective listening tests. The intrusive short-time objective speech intelligibility (STOI) metric [24] was used because it is widely used to evaluate algorithm performance, but to our knowledge has not been tested with NELE algorithms in noise and reverberation before. In addition, a measure to predict perceived listening effort based on automatic speech recognition was used, which was shown to be highly correlated with experimental data for AdaptDRC-processed speech in noise [25], but so far was not tested for NELE-processed speech in reverberation either. Since this measure includes a mapping to the same perceptual scale as used in the experiment, it can be quantitatively compared to the subjective data.

The subjective evaluations were conducted with 17 listeners with audiotically normal hearing, who listened to the stimuli via headphones. The task was to rate the perceived listening effort associated with understanding the target talker on a categorical 13-point scale ranging from "no effort" (1 Effort Scaling Categorical Unit, ESCU) to "extreme effort" (13 ESCU) [20] on a graphical user interface. Target stimuli were random sentences from the German Oldenburg sentence test [17], which had been played back and recorded using a KEMAR dummy head in a room with electronically adjustable reverberation times. The reverberant speech level was always fixed at 65 dB SPL at the dummy head position, and cafeteria noise [22] was scaled and played back in the same room to produce SNRs of  $-5$  and  $10$  dB. For each SNR, three different reverberation times were used, corresponding to simulations of an opera hall ( $T_{30} = 1.08$  s), a cafeteria ( $T_{30} = 2.40$  s), and a train station ( $T_{30} = 4.14$  s). Each of these six conditions was measured with

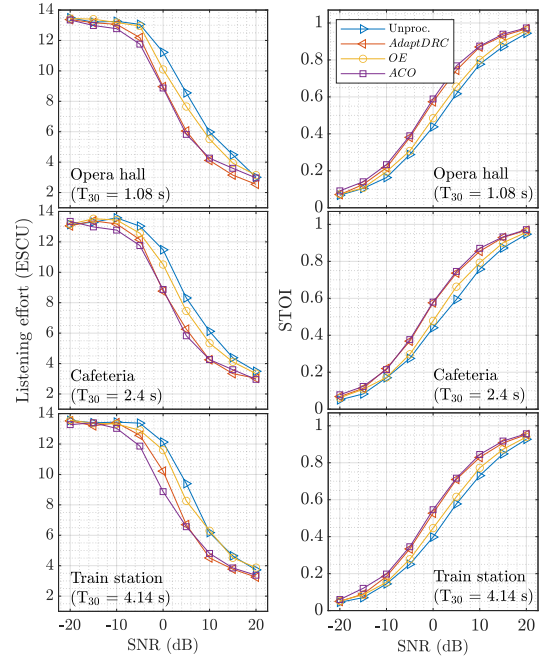


Figure 3: Results of the instrumental evaluation for the listening effort model (left) and STOI (right). Each panel presents the computed measures as functions of the SNR for the different processing types.

unprocessed speech as well as with speech processed by the three algorithms (AdaptDRC, OE, and ACO). The ratings for the four signals in each condition were carried out at the same time using a method motivated by MUSHRA tests, i.e., listeners could switch between the test signals, a clean-speech reference and a low-quality anchor, and both rate and rank the perceived listening effort. The six conditions as well as the processing types within a condition were randomized in their order. Non-parametric Friedman tests followed by pair-wise Wilcoxon rank sum tests were conducted separately for each combination of reverberation and SNR to assess if differences between algorithms were statistically significant. Post-hoc tests were Bonferroni-corrected for multiple comparisons.

For the instrumental evaluations, ten sentences from the German Oldenburg sentence test were used. The speech was first processed by each processing type and convolved with the RIRs as in the experiment. The reverberant speech was then summed with the same noise type and scaled for SNRs between  $-20$  and  $20$  dB.

#### 3.2. Results

Figure 3 shows the results of the instrumental evaluation for the listening effort model (left, lower is better) and STOI (right, higher is better). Each panel presents the computed measures as functions of SNR for the different processing types. In all conditions, the measures showed the expected improvement with increasing SNR. In comparison to unprocessed speech (right-pointing triangles), all algorithms produced an improvement over the entire SNR range with the exception of ceiling and floor effects at very high and low SNRs, respectively. The improvement depended on processing type, and was largest for the ACO algorithm (squares), although the instrumental measures

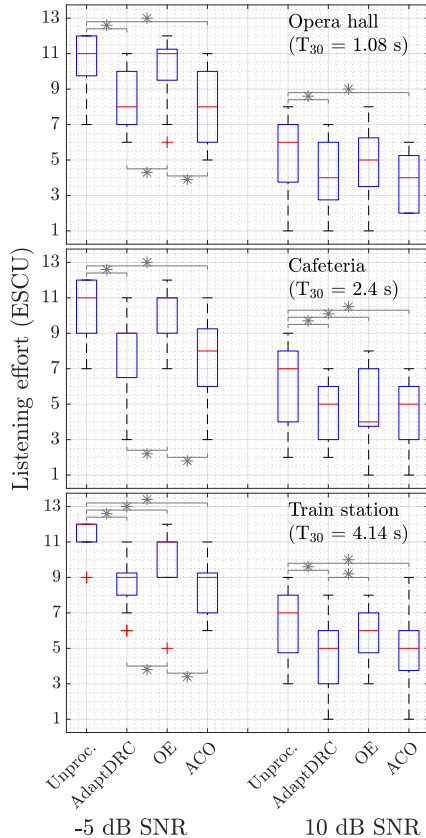


Figure 4: Measured listening effort ratings. Statistically significant differences between processing types are marked by an asterisk.

indicated a very similar improvement for the AdaptDRC algorithm alone (left-pointing triangles). For these two algorithms, the improvement corresponded to SNR shifts of the psychometric functions of about 5 dB (listening effort measure) and 4 dB (STOI), as indicated by the horizontal distance between the functions in the mid-range of the psychometric functions. In contrast, the benefit obtained by the OE algorithm alone (circles) was smaller, corresponding to about 2 dB SNR. The algorithm benefit did not appear to differ much between the different reverberation conditions.

Experimentally measured listening effort ratings are shown as boxplots in Figure 4. In general, the results were in line with what could be expected from the instrumental measures. In particular, perceived listening effort was significantly lower for speech processed by ACO and AdaptDRC compared to unprocessed speech (benefit about 2-3 ESCU when comparing median values). For the OE algorithm, the benefit was absent (at -5 dB SNR) or present but smaller (at 10 dB SNR), reaching statistical significance only in the cafeteria condition. In general, the observed algorithm benefit was slightly larger at -5 than at 10 dB SNR, and similar for all reverberation conditions.

#### 4. Discussion

The comparison of ACO (i.e., the combination of AdaptDRC and OE) and AdaptDRC indicated no consistent advantage of the implemented combination. For some conditions, the instru-

mental measures and (by trend) the experimental data indicated a benefit, while for other conditions ACO did not produce better results than the AdaptDRC algorithm alone. This is an interesting outcome because evaluations within the recent Hurricane Challenge [26] suggest that ACO showed better performance in highly reverberant conditions compared to conditions with less reverberation. This would not necessarily be expected for AdaptDRC processing alone, but likely represents the effects of the OE-processing which reduces the self-masking of speech and is, hence, particularly effective in more reverberant conditions [14]. However, a direct comparison between the present data and the Challenge outcome is difficult because the isolated AdaptDRC algorithm did not participate in the Challenge. Further research seems necessary to investigate the specific contributions of the AdaptDRC- and OE-components as well as to investigate the dependency of the algorithm benefit on acoustical conditions. Conversely, the present data indicate that the AdaptDRC algorithm is applicable also in reverberant conditions despite the fact that it was not designed for that purpose. The benefit measured in this study (4-5 dB SNR) was comparable to the benefit measured in a previous study for anechoic speech in a comparable cafeteria noise [7].

Another important outcome of this study was that the employed listening effort measure produced very accurate predictions of the experimental data. While this was already shown for AdaptDRC-processed (anechoic) speech and different masking noises [25], the present study indicates that the measure generalizes well across different types of NELE algorithms as well as reverberant conditions. Importantly, it agreed well with the data also with respect to the amount (or absence) of a benefit in terms of perceived listening effort compared to unprocessed speech. This makes the measure a potentially powerful tool for algorithm comparisons and parameter tuning in realistic listening conditions. Because it is non-intrusive it is generally possible to employ the measure as an online monitoring system in the future, although a non-realtime version of the measure was used in this study. This would further increase the measure's practical use, e.g., as a control instance to steer algorithm parameters in time-varying acoustic conditions.

#### 5. Conclusions

The following conclusions can be drawn from this study:

1. The AdaptDRC algorithm developed for speech in noise [7] also performs well in reverberant conditions.
2. The combination of the AdaptDRC algorithm with the reverberation-dependent algorithm OE [14] produces only small additional improvements compared to AdaptDRC alone for the tested conditions.
3. All of the experimentally observed trends are well captured by the tested instrumental measures. Specifically, the single-ended listening effort measure [25] yields quantitatively accurate predictions, indicating its general applicability as a tool for evaluating NELE algorithms.

#### 6. Acknowledgements

The authors would like to thank Julian Grosse and Steven van de Par for providing the algorithm OE and their support for this study. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 352015383 – SFB 1330 A1/C1, Projektnummer 390895286 - EXC 2177/1, and grant no. RE 4160/1-1.

## 7. References

- [1] M. Morimoto, H. Sato, and M. Kobayashi, "Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces," *J. Acoust. Soc. Am.*, vol. 116, no. 3, pp. 1607–1613, Sep. 2004.
- [2] J. Rennies, H. Schepker, I. Holube, and B. Kollmeier, "Listening effort and speech intelligibility in listening situations affected by noise and reverberation," *J. Acoust. Soc. Am.*, vol. 136, no. 5, pp. 2642–2653.
- [3] B. Sauert and P. Vary, "Near end listening enhancement optimized with respect to speech intelligibility index and audio power limitations," in *Proc. Europ. Signal Proc. Conf. (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 1919–1923.
- [4] M. Cooke, C. Mayo, and C. Valentini-Botinhao, "Intelligibility enhancing speech modifications: The Hurricane Challenge," in *Proc. Interspeech*, Lyon, France, Aug. 2013, pp. 3552–3556.
- [5] C. H. Taal, R. C. Hendriks, and R. Heusdens, "Speech Energy Redistribution for Intelligibility Improvement in Noise Based on a Perceptual Distortion Measure," *Computer Speech & Language*, vol. 28, no. 4, pp. 858–872, July 2014.
- [6] W. B. Kleijn, J. B. Crespo, R. C. Hendriks, P. N. Petkov, B. Sauert, and P. Vary, "Optimizing speech intelligibility in a noisy environment," *IEEE Signal Proc. Mag.*, vol. 32, no. 2, Mar. 2015, pp. 43–54.
- [7] H. Schepker, J. Rennies, and S. Doclo, "Speech-in-noise enhancement using amplification and dynamic range compression controlled by the speech intelligibility index," *J. Acoust. Soc. Am.*, vol. 138, no. 5, pp. 2692–2706, Nov. 2015.
- [8] T.-C. Zorila, V. Kandia, and Y. Stylianou, "Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression," in *Proc. Interspeech*, Portland, USA, Sept. 2012, pp. 635–638.
- [9] M. Niermann, C. Thierfeld, P. Jax, and P. Vary, "Time Domain Approach for Listening Enhancement in Noisy Environments," in *Proc. ITG. Conf. Speech Comm.*, Paderborn, Germany, Oct. 2016, pp. 282–286.
- [10] A. Mertins, T. Mei, and M. Kallinger, "Room impulse response shortening/reshaping with infinity and p-norm optimization," *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, no. 2, pp. 249–259, Feb. 2010.
- [11] A. Kusumoto, T. Arai, K. Kinoshita, N. Hodoshima, and N. Vaughan, "Modulation enhancement of speech by a preprocessing algorithm for improving intelligibility in reverberant environments," *Speech Commun.*, vol. 45, no. 2, pp. 101–113, Feb. 2005.
- [12] T. Arai, N. Hodoshima, and K. Yasu, "Using steady-state suppression to improve speech intelligibility in reverberant environments for elderly listeners," *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, no. 7, pp. 1775–1780, Sep. 2010.
- [13] P. N. Petkov and Y. Stylianou, "Adaptive Gain Control for Enhanced Speech Intelligibility under Reverberation," *IEEE Sig. Proc. Letters*, vol. 23, no. 10, pp. 1434–1438, 2016.
- [14] J. Grosse and S. van de Par, "A Speech Preprocessing Method Based on Overlap-Masking Reduction to Increase Intelligibility in Reverberant Environments," *J. Audio Eng. Soc.*, vol. 65, no. 1/2, pp. 31–41, Jan. 2017.
- [15] H. Schepker, D. Hülsmeyer, J. Rennies, S. Doclo, "Model-based integration of reverberation for noise-adaptive near-end listening enhancement," in *Proc. Interspeech*, Dresden, Germany, Aug. 2015, pp. 75–79.
- [16] R. C. Hendriks, J. B. Crespo, J. Jensen, and C. H. Taal, "Optimal Near-End Speech Intelligibility Improvement Incorporating Additive Noise and Late Reverberation Under an Approximation of the Short-Time SII," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 5, pp. 851–862, May 2015.
- [17] K. Wagener, V. Kühnel, and B. Kollmeier, "Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests (Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test)," *Z. Audiol.*, vol. 38, pp. 4–15, Sep. 1999.
- [18] K. Smeds, F. Wolters and M. Rung, "Estimation of signal-to-noise ratios in realistic sound scenarios," *J. Am. Acad. Aud.*, vol. 26, no. 2, pp. 183–196, Feb. 2015.
- [19] J. Rennies, A. Pusch, H. Schepker, and S. Doclo, "Evaluation of a near-end listening enhancement algorithm by combined speech intelligibility and listening effort measurements," in *J. Acoust. Soc. Am.*, vol. 144, no. 4, EL315–EL321, Oct. 2018.
- [20] M. Krueger, M. Schulte, M. Zokoll, K. Wagener, M. Meis, T. Brand, and I. Holube, "Relation between listening effort and speech intelligibility in noise," *Am. J. Audiol.*, vol. 26, pp. 378–392, Oct. 2017.
- [21] E. Haensler and G. Schmidt, "Speech and audio processing in adverse environments," Springer Verlag, Berlin, Germany, 2008.
- [22] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP Journal on Advances in Signal Processing*, vol. 6, 2009.
- [23] H. Schepker, K. Haeder, J. Rennies, and I. Holube, "Perceived listening effort and speech intelligibility in reverberation and noise for hearing-impaired listeners," *Int. J. Audiol.*, vol. 55, no. 12, Dec. 2016, pp. 738–747.
- [24] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, 2010, pp. 4214–4217.
- [25] R. Huber, A. Pusch, N. Moritz, J. Rennies, H. Schepker, and B.T. Meyer, "Objective assessment of a speech enhancement scheme with an automatic speech recognition-based system", in *Speech Communication; 13th ITG-Symposium*, Oldenburg, Germany, 2018, pp. 1–5.
- [26] J. Rennies, H. Schepker, C. Valentini-Botinhao, and M. Cooke, "Intelligibility-enhancing speech modifications – The Hurricane Challenge 2.0," *Proc. Interspeech*, Shanghai, China, Oct. 2020