



Model-based integration of reverberation for noise-adaptive near-end listening enhancement

Henning Schepker¹, David Hülsmeier², Jan RENNIES², Simon Doclo^{1,2}

¹University of Oldenburg, Signal Processing Group, Dept. of Medical Physics and Acoustics and the Cluster of Excellence Hearing4All, Oldenburg, Germany

²Fraunhofer IDMT, Project Group Hearing, Speech, and Audio Technology, Oldenburg, Germany

{henning.schepker, simon.doclo}@uni-oldenburg.de,
{david.huelsmeier, jan.rennies}@idmt.fraunhofer.de

Abstract

Speech intelligibility is an important factor for successful speech communication in today’s society. So-called near-end listening enhancement (NELE) algorithms aim at improving speech intelligibility in conditions where the (clean) speech signal is accessible and can be modified prior to its presentation. However, many of these algorithms only consider the detrimental effect of noise and disregard the effect of reverberation. Therefore, in this paper we propose to additionally incorporate the detrimental effects of reverberation into noise-adaptive near-end listening enhancement algorithms. Based on the Speech Transmission Index (STI), which is widely used for speech intelligibility prediction, the effect of reverberation is effectively accounted for as an additional noise power term. This combined noise power term is used in a state-of-the-art noise-adaptive NELE algorithm. Simulations using two objective measures, the STI and the short-time objective intelligibility (STOI) measure demonstrate the potential of the proposed approach to improve the predicted speech intelligibility in noisy and reverberant conditions.

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

1. Introduction

In many speech communication devices, e.g., public address systems and mobile telephony, a high quality of communication needs to be provided. To maintain a high quality of communication, a high speech intelligibility is an important factor. However, in many situations the speech signal is (severely) degraded by noise and/or reverberation, resulting in reduced speech intelligibility and increased listening effort [1, 2]. A simple, yet effective approach to maintain a good speech intelligibility is to raise the speech level prior to its presentation and hence increase the signal-to-noise ratio (SNR). Although being also found in human speech production as the so-called Lombard effect [3] and hence being an attractive modification, this approach may lead to an overload of the amplification system and unpleasantly high sound levels. Consequently, approaches that increase speech intelligibility while maintaining equal powers of the unprocessed and processed speech signal are desirable.

Many near-end listening enhancement (NELE) algorithms have been proposed to increase speech intelligibility in noise under equal power constraints [4, 5, 6, 7, 8, 9, 10, 11, 12].

This work was supported in part by the Research Unit FOR 1732 “Individualized Hearing Acoustics” and the Cluster of Excellence 1077 “Hearing4All”, funded by the German Research Foundation (DFG).

These include algorithms that apply only spectral changes to the speech signals [6, 7], dynamic range compression (DRC) [8] or combined approaches [9, 10]. However, these algorithms usually only consider the detrimental effect of noise on speech intelligibility and neglect the presence of reverberation. On the contrary, algorithms that aim to increase speech intelligibility in reverberant environments, e.g., listening room compensation algorithms (LRC) [13] and modulation enhancing algorithms [14], mainly neglect the presence of noise. One of the first attempts to consider noise as well as reverberation in the design of pre-processing algorithms to increase speech intelligibility was provided in [11, 12]. The approach in [12] assumes the room impulse response (RIR) to be an exact exponential decaying function. Although being a convenient assumption in their mathematical framework, this does not necessarily hold for realistic (e.g., measured) RIRs. Therefore, in this paper we present an approach that can make use of the full RIR information and allows to incorporate reverberation information into noise-adaptive NELE algorithms. Based on the concept of the Speech Transmission Index (STI) [15, 16] the influence of reverberation is incorporated as an additional noise term, making it applicable, in principle, to any noise-adaptive NELE algorithm. Hence, in contrast to LRC algorithm which usually aim at equalizing the RIR mainly spectral information of reverberation is used. Furthermore, the proposed approach can be used even with only limited knowledge, e.g., the broadband reverberation time T_{60} , where it assumes an exponentially decaying function, similar to [12].

This paper is organized as follows. In Section 2 the considered scenario and some definitions are provided. In Section 3 the novel model-based integration of reverberation is derived. In Section 4 the concept of *AdaptDRC* algorithm from [9] is briefly reviewed and is extended to the noisy and reverberant case referred to as *AdaptDRCrev*. Experimental results using two different noises and simulated as well as measured RIRs are presented in Section 5 that demonstrate the potential of the proposed approach.

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 and reverberant speech signal $y[k]$, which is the mixture of the modified speech signal $\tilde{s}[k]$ convolved with the RIR $h[k]$ between the loudspeaker and the

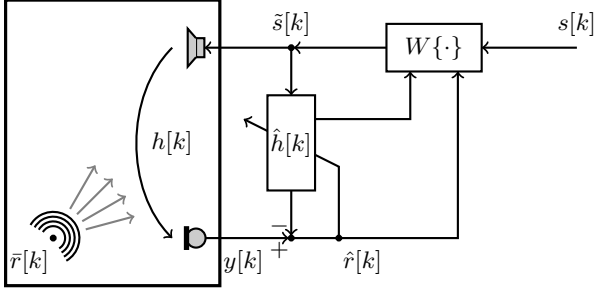


Figure 1: Considered acoustical scenario.

microphone and the additive noise disturbance $r[k]$, i.e.,

$$y[k] = \tilde{s}[k] * h[k] + \underbrace{\tilde{r}[k] * g[k]}_{r[k]}, \quad (1)$$

where $*$ denotes convolution. Additionally it is assumed that the noise signal $r[k]$ can be modeled as the convolution of the noise source signal $\tilde{r}[k]$ and the RIR $g[k]$ between the noise source and the microphone. An estimate $\hat{r}[k]$ of the noise signal $r[k]$ as well as an estimate $\hat{h}[k]$ of the RIR $h[k]$ between the loudspeaker and the microphone can be obtained by using, e.g., adaptive filtering techniques to model $h[k]$ [17]. 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 speech signal $\tilde{s}[k]$ is computed as

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

In the following perfect knowledge of the RIR is assumed, i.e., $\hat{h}[k] = h[k]$ and hence a perfect noise estimate is available, i.e., $\hat{r}[k] = r[k]$. The goal of NELE algorithms commonly is to find a weighting function $W\{\cdot\}$ that leads to an improved speech intelligibility of $\tilde{s}[k] + r[k]$ compared to $s[k] + r[k]$. In contrast, the goal here is to find a weighting function $W\{\cdot\}$ that leads to an improved speech intelligibility of $\tilde{s}[k] * h[k] + r[k]$ compared to $s[k] * h[k] + r[k]$. In addition, to avoid trivial broadband amplification an equal power constraint is imposed on $\tilde{s}[k]$.

The following processing framework is applied. The speech signal $s[k]$, the estimated noise signal $\hat{r}[k]$ and the RIR $\hat{h}[k]$ are split into N subband signals $s_n[k]$, $\hat{r}_n[k]$ and $\hat{h}_n[k]$, $n = 1, \dots, N$ using a real-valued filter bank. In the implementation an all-pass filterbank based on doubly-complementary IIR filters is used [18] to split the signals into $N = 8$ octave bands with center frequencies from 125 Hz to 16 kHz. Each subband signal is framed into non-overlapping blocks of length M , i.e., $s_n^l[m] = s_n[lM + m]$ and $\hat{r}_n^l[m] = \hat{r}_n[lM + m]$, $m = 0, \dots, M - 1$ with block index l . The speech power in l -th block in the n -th subband is equal to

$$\phi_s[n, l] = \frac{1}{M} \sum_{m=0}^{M-1} (s_n^l[m])^2 \quad (3)$$

Similarly the noise source power $\phi_{\tilde{r}}[n, l]$ and the estimated noise power $\phi_{\hat{r}}[n, l]$ in l -th block in the n -th subband are equal to

$$\phi_{\tilde{r}}[n, l] = \frac{1}{M} \sum_{m=0}^{M-1} (\tilde{r}_n^l[m])^2 \quad (4)$$

$$\phi_{\hat{r}}[n, l] = \frac{1}{M} \sum_{m=0}^{M-1} (\hat{r}_n^l[m])^2 \quad (5)$$

where $\tilde{r}_n^l[m] = \tilde{r}_n[lM + m]$ is the noise source signal $\tilde{r}[k]$ in l -th block in the n -th subband.

3. Model-based integration of reverberation

To consider reverberation in noise-adaptive NELE algorithms the goal in this section is to derive a model-based so-called apparent noise power based on the STI [16] that considers both noise and reverberation. It is shown that based on the concept of the modulation transfer function (MTF) employed in the calculation of the STI a single noise power term in each subband can be computed that accounts for both the noise and reverberation.

The STI is based on the observation that a reduction in modulation depth due to additive noises and reverberation is highly correlated with speech intelligibility [16]. For the calculation of the STI the following signal model is assumed

$$y[k] = (s[k] + \tilde{r}[k]) * h[k] \quad (6)$$

Note that this signal model is different from the signal model in (1) which is assumed in the design of NELE algorithms. Despite this difference in the signal models, in the following it is shown how reverberation can be accounted for in noise-adaptive NELE algorithms based on the concept of the STI.

The STI calculation rules provide analytic relationships between the SNR and the modulation index $m_{noise}[n, f]$ as well as the reverberation time $T_{60}[n]$ in the n -th subband and the modulation index $m_{reverb}[n, f]$, where f denotes the modulation frequency. For the sake of clarity the dependency of the speech signal power and the noise signal powers on the block index l is omitted.

The modulation index for influences of noise is equal to

$$m_{noise}[n, f] = \frac{1}{1 + \frac{\phi_{\tilde{r}}[n]}{\phi_s[n]}}. \quad (7)$$

Note that $m_{noise}[n, f]$ is actually independent of the modulation frequency f . The modulation index for the influence reverberation (assuming a perfect exponential decay) is equal to [15]

$$m_{reverb}[n, f] = \frac{1}{\sqrt{1 + (2\pi f \frac{T_{60}[n]}{13.8})^2}}, \quad (8)$$

where $T_{60}[n]$ is the reverberation time in the n -th subband. Alternatively, assuming knowledge of the true RIR $h[k]$ of length L_h the modulation index $m_{reverb}[n, f]$ can be computed as [19]

$$m_{reverb}[n, f] = \frac{\sum_{k=0}^{L_h-1} e^{-j2\pi f k / f_s} h_n^2[k]}{\sum_{k=0}^{L_h-1} h_n^2[k]}, \quad (9)$$

where f_s is the sampling frequency.

The combined influence of noise and reverberation is then modeled by the multiplication of $m_{noise}[n]$ in (7) and $m_{reverb}[n, f]$ in (8) (assuming a perfect exponential decay) or $m_{reverb}[n, f]$ in (9) (assuming knowledge about the true RIR) yielding the modulation index

$$m[n, f] = \frac{1}{1 + \frac{\phi_{\tilde{r}}[n]}{\phi_s[n]}} \cdot m_{reverb}[n, f]. \quad (10)$$

The modulation index $m[n, f]$ can then be used to calculate a so-called apparent SNR which effectively accounts for both

noise and reverberation, i.e.,

$$\frac{\phi_s[n]}{\phi_r^{app}[n, f]} = \frac{m[n, f]}{1 - m[n, f]}, \quad (11)$$

where $\phi_r^{app}[n, f]$ is the apparent noise power which effectively takes into account both the noise and reverberation. The apparent noise power can then be used in a noise-adaptive NELE algorithms instead of using only the noise power.

Based on the concept of the apparent SNR and using (7) as a function of the apparent noise power, (10) can be reformulated as

$$\frac{1}{1 + \frac{\phi_r^{app}[n, f]}{\phi_s[n]}} = \frac{1}{1 + \frac{\phi_r[n]}{\phi_s[n]}} \cdot m_{reverb}[n, f] \quad (12)$$

solving for the apparent noise power $\phi_r^{app}[n, f]$ yields

$$\boxed{\phi_r^{app}[n, f] = \phi_s[n] \left(\frac{1}{m_{reverb}[n, f]} - 1 \right) + \phi_r[n] \frac{1}{m_{reverb}[n, f]}} \quad (13)$$

Thus the apparent noise power is the sum of the reverberant part of the speech power (first term in (13)) and the reverberant noise power (the second term in (13)).

Following the calculation rules of the STI [15, 16] an average across modulation frequencies of the apparent noise power is calculated in the log-domain, i.e.,

$$10 \log_{10} \bar{\phi}_r^{app}[n] = \frac{1}{|\mathcal{F}|} \sum_{f \in \mathcal{F}} 10 \log_{10} \phi_r^{app}[n, f]. \quad (14)$$

where \mathcal{F} is the set of 14 modulation frequencies used in the STI [16] and $|\mathcal{F}|$ is its cardinality. The average apparent noise power $\bar{\phi}_r^{app}[n]$ can then be used in noise-adaptive NELE algorithms to account for the combined effect of noise and reverberation instead of the estimated noise power $\phi_r[n]$ which only accounts for the noise.

Unfortunately, the apparent noise power in (13) requires knowledge of the noise source power $\phi_r[n]$ which is usually not accessible in practice. In order to apply $\phi_r^{app}[n, f]$ in NELE algorithms, it is helpful to consider the difference in the signal models assumed for NELE algorithms in (1) and the signal model assumed for the calculation of the STI in (6). The last term in (13) clearly corresponds to a modified, i.e., reverberated, version of the power of the noise source signal. However, the modulation index $m[n, f]$ accounts only for the reverberation introduced by the RIR $h[k]$ between the loudspeaker and the microphone, while the RIR $g[k]$ between the noise source and the microphone will generally differ substantially from $h[k]$. Furthermore, in practice the noise source signal $\bar{r}[k]$ is not accessible but only access to the estimated noise $\hat{r}[k]$ of the reverberant noise signal $r[k] = \bar{r}[k] * g[k]$ is available. Thus to allow for the practical application, based on the above reasoning it is proposed to substitute the last term in (13) by the estimated noise power $\phi_{\hat{r}}[n]$, yielding

$$\boxed{\phi_{\hat{r}}^{app}[n, f] = \phi_{s,n} \left(\frac{1}{m_{reverb}[n, f]} - 1 \right) + \phi_{\hat{r}}[n]} \quad (15)$$

which is used in (14) to compute the average apparent noise power $\bar{\phi}_{\hat{r}}^{app}[n]$ instead of the definition in (13).

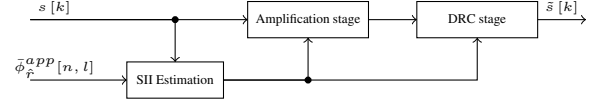


Figure 2: Schematic flow-graph of the *AdaptDRCrev* algorithm

4. AdaptDRCrev Algorithm

The *AdaptDRC* algorithms [9] aims at improving speech intelligibility in noisy environments. A brief description of the algorithm is provided below and the reader is referred to [9] for a more detailed description of the *AdaptDRC* algorithm. By using the proposed model-based integration of reverberation in Section 3 the *AdaptDRC* algorithm is extended for the noisy and reverberant case, hence referred to as *AdaptDRCrev* in the remainder. A schematic overview of the *AdaptDRCrev* algorithm is provided in Figure 2. The *AdaptDRC* and *AdaptDRCrev* algorithms combine a time- and frequency-dependent amplification stage with a time- and frequency-dependent dynamic range compression stage that are both a function of the estimated Speech Intelligibility Index (SII) [20] which depends on the speech and (apparent) noise power. Both stages aim to modify the speech signal only in case of low predicted speech intelligibility, while in case of high predicted speech intelligibility the speech signal remains unmodified.

The time- and frequency-dependent amplification stage aims at increasing the speech power in high frequency regions in case of low predicted speech intelligibility by applying the following gain function

$$w[n, l] = \sqrt{\frac{(\phi_s[n, l])^{SII[l]} \cdot \sum_{\lambda=1}^N \phi_s[\lambda, l]}{\sum_{\lambda=1}^N (\phi_s[\lambda, l])^{SII[l]} \cdot \phi_s[n, l]}} \quad (16)$$

with $\hat{SII}[l]$ an estimate of the SII in the l -th block. Hence in case of low predicted speech intelligibility, i.e., $\hat{SII}[l] \rightarrow 0$,

$w[n, l] = \sqrt{\frac{\sum_{\lambda=1}^N \phi_s[\lambda, l]}{\frac{1}{N} \sum_{\lambda=1}^N \phi_s[\lambda, l]}}$ resulting in a uniform distribution of the speech power across all N subbands, while for high predicted speech intelligibility, i.e., $\hat{SII}[l] \rightarrow 1$, $w[n, l] = 1$ and hence no modification is applied.

The dynamic range compression stage aims at amplifying low-levels signals and attenuating high-levels signals that are assumed to be well audible by applying a time-dependent non-linear gain in each subband, i.e.,

$$s_n^l[m] = s_n^l[m] \cdot p(\lambda_n[l], (s_n^l[m])^2), m = 0, \dots, M - 1 \quad (17)$$

where $p(\lambda_n[l], (s_n^l[m])^2)$ is the non-linear gain function that depends on the time- and subband-dependent input-output-characteristic $\lambda_n[l]$ of the DRC stage and the estimated envelope of the speech signal $s_n^l[m]$ [9]. While in the *AdaptDRC* algorithm $\hat{SII}[l]$ and $\lambda_n[l]$ depend on the estimated noise power $\phi_{\hat{r}}[n]$, in the *AdaptDRCrev* the apparent noise power $\bar{\phi}_{\hat{r}}^{app}[n]$ according to (14) is used instead effectively accounting for the combined effect of noise and reverberation.

5. Evaluation

In this section the proposed model-based integration of reverberation is evaluated. Specifically the new *AdaptDRCrev* algorithm as introduced in Section 4 which considers noise and reverberation is compared to the *AdaptDRC* algorithm which considers only noise. Evaluations were performed for both the

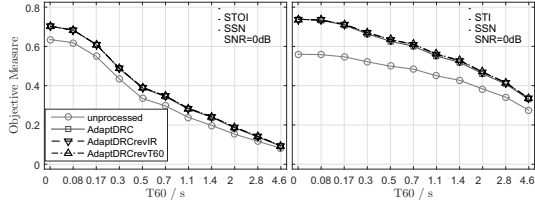


Figure 3: Results for STOI (left panel) and STI (right panel) as a function of the reverberation time T_{60} using simulated RIRs and the SSN for a fixed SNR of 0 dB for the different processing conditions.

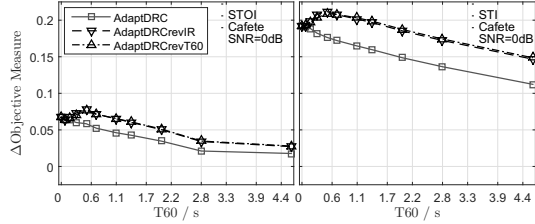


Figure 4: Improvements compared to the unprocessed signal for STOI (left panel) and STI (right panel) as a function of the reverberation time T_{60} using simulated RIRs and the non-stationary cafeteria noise for a fixed SNR of 0 dB for the different processing conditions.

modulation index calculated using a frequency-independent reverberation time $T_{60}[n] = T_{60}$ according to (8) (*AdaptDRCrevT60*) as well as the modulation index using the RIR information according to (9) (*AdaptDRCrevIR*). Two objective measures were used that have shown high correlations with speech intelligibility in previous studies: the STI [16] and the short-term objective intelligibility (STOI) measure [21]. Speech material was taken from the Oldenburg Sentence Test recorded by one male German speaker [22]. Ten sentences were randomly selected from the corpus. Two different noises were used, a stationary speech-shaped noise (SSN) and non-stationary cafeteria noise. Simulated RIRs using the image method [23] and measured RIRs were used to evaluate the impact of reverberation. To achieve different degrees of reverberation for the simulated RIRs the room dimensions were kept fixed (6 m \times 8 m \times 2.5 m) and the absorption coefficient was varied. All signals were sampled using a sampling frequency of $f_s = 44.1$ kHz.

Figure 3 depicts the results for the SSN and a fixed SNR of 0 dB for different reverberation times T_{60} of the simulated RIRs. As can be observed all three algorithms improve over the unprocessed condition. Furthermore, both *AdaptDRCrev* algorithms improves slightly over *AdaptDRC* as can be seen in the STI measure, while this is not observed for the STOI measure. To see the impact of incorporating reverberation more clearly in the following the improvement compared to the unprocessed signal is shown.

Figure 4 depicts the improvement for the cafeteria noise and a fixed SNR of 0 dB for different reverberation times T_{60} of the simulated RIRs. Again all three algorithms improve over the unprocessed condition. Both *AdaptDRCrev* algorithms improve over the *AdaptDRC* algorithm demonstrating the benefit of considering reverberation in noise-adaptive NELE algorithms.

Figures 5 and 6 show the improvement for a fixed RIR measured in a conference room with a reverberation time $T_{60} = 0.6$ s as a function of the SNR for the SSN and the cafeteria noise, respectively. All algorithms improvement over the unprocessed condition. For the SSN only marginal improvements of both *AdaptDRCrev* algorithms over the *AdaptDRC* algorithm

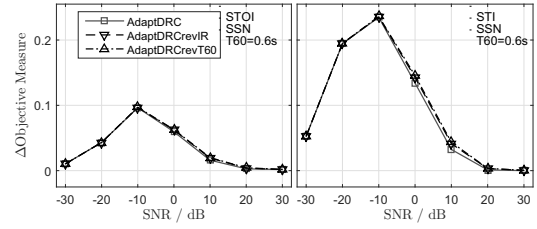


Figure 5: Improvements compared to the unprocessed signal for STOI (left panel) and STI (right panel) as a function of the SNR for the different processing conditions in the SSN using a measured RIR from a conference room with $T_{60} = 0.6$ s.

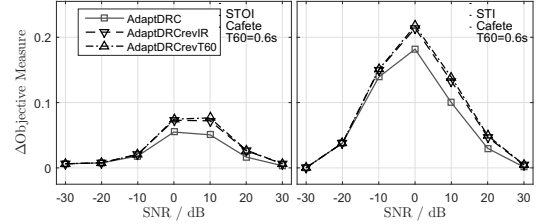


Figure 6: Improvements compared to the unprocessed signal for STOI (left panel) and STI (right panel) as a function of the SNR for the different processing conditions in the non-stationary cafeteria noise using a measured RIR from a conference room with $T_{60} = 0.6$ s.

are observed. However, for the non-stationary cafeteria noise improvements for SNRs > 0 dB are visible.

In general these results demonstrate the potential of incorporating reverberation information in noise-adaptive NELE algorithms to increase the performance in noisy and reverberant conditions. Even when using only broadband T_{60} information, i.e., when using (8) improvements are in the same range compared to using the complete RIR in (9). These improvements, however, appear to depend on the noise characteristics, i.e., their spectral overlap with the speech signal as well as their temporal characteristics. Furthermore, predicted speech intelligibility improvements are largest for longer reverberation times. The additional improvement achieved by incorporating reverberation information are smaller compared to the improvements obtained by considering only noise. This is in line with results from previous studies [12], where predictions of the STI appear to even indicate improvements of using only noise information over using noise and reverberation information.

6. Conclusions

In this paper a new approach to incorporate reverberation information into noise-adaptive NELE algorithms aiming to increase speech intelligibility in noisy and reverberant environments has been proposed. Using the concept of the STI reverberation is incorporated as an additional noise term that can be used, in principle, with any noise-adaptive NELE algorithm. The proposed approach was incorporated into the *AdaptDRC* algorithm. Experimental results using the new *AdaptDRCrev* algorithm demonstrate the potential of the proposed method to increase predicted speech intelligibility. However, the improvements of the *AdaptDRCrev* algorithm over the *AdaptDRC* algorithm appear to be small compared to the improvements that were observed for the *AdaptDRC* algorithm compared to the unprocessed signals. To which extent these results reflect a noticeable perceptual benefit for listeners in reverberant conditions has to be evaluated in formal listening tests and is left for future work.

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, 2004.
- [2] J. Rannies, 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, Nov. 2014.
- [3] W. Van Summers, D. B. Pisoni, R. H. Bernacki, R. I. Pedlow and M. A. Stokes, "Effects of noise on speech production: Acoustic and perceptual analyses", *J. Acoust. Soc. Am.*, vol. 84, no. 3, pp. 917–928, 1988.
- [4] M. Cooke, C. Mayo, and C. Valentini-Botinhao, "Intelligibility-enhancing speech modifications: the Hurricane challenge", *Proc. Interspeech*, Lyon, France, Aug. 2013, pp. 3552–3556.
- [5] 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, 2015, pp. 43–54.
- [6] B. Sauert and P. Vary, "Near end listening enhancement optimized with respect to speech intelligibility index and audio power limitations", *Proc. Europ. Signal Proc. Conf. (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 1919–1923.
- [7] 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, 2014.
- [8] R. Niederjohn and J. Grotelueschen, "The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression", *IEEE Trans. Acoust. Speech*, vol. 24, no. 4, pp. 277–282, 1976.
- [9] H. Schepker, J. Rannies, and S. Doclo, "Improving speech intelligibility in noise by SII-dependent preprocessing using frequency-dependent amplification and dynamic range compression", *Proc. Interspeech*, Lyon, France, Aug. 2013, pp. 3577–3581.
- [10] T.-C. Zorila, V. Kandia, and Y. Stylianou, "Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression", *Proc. Interspeech*, Portland, USA, Sept. 2012, pp. 635–638.
- [11] J. B. Crespo and R. C. Hendriks, "Speech reinforcement in noisy reverberant environments using a perceptual distortion measure", *Proc. Int. Conf. Acoust. Speech Sign. Process. (ICASSP)*, Firenze, Italy, May 2014, pp. 910–914.
- [12] J. B. Crespo and R. C. Hendriks, "Speech reinforcement with a globally optimized perceptual distortion measure for noisy reverberant channels", *Proc. Int. Workshop Acoust. Sign. Enhance. (IWAENC)*, Antibes - Juan Les Pins, France, Sept. 2014, pp. 90–94.
- [13] 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.
- [14] 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, 2005.
- [15] T. Houtgast and H. J. M. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria", *J. Acoust. Soc. Am.*, vol. 77, no. 3, pp. 1069–1077, Mar. 1985.
- [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] E. Haensler and G. Schmidt, "Speech and audio processing in adverse environments", Springer Verlag, Berlin, Germany, 2008.
- [18] P. Regalia, S. Mitra, P. Vaidyanathan, M. Renfors, and Y. Neuvo, "Tree-structured complementary filter banks using all-pass sections", *IEEE Trans. Circuits Syst.*, vol. 34, no. 12, 1470–1484, 1987.
- [19] M. R. Schroeder, "Modulation transfer functions: Definition and measurement", *Acustica*, vol. 49, pp. 179–182, 1981.
- [20] ANSI, "Methods for calculation of the speech intelligibility index", American National Standard ANSI S3.5-1997, New York, USA, 1997.
- [21] C. Taal, R. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time frequency weighted noisy speech", *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, 2011.
- [22] 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, 1999.
- [23] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics", *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.