SPEECH REINFORCEMENT IN NOISY REVERBERANT CONDITIONS UNDER AN APPROXIMATION OF THE SHORT-TIME SII

Richard C. Hendriks[1], *João B. Crespo, Jesper Jensen*[2,3] *and Cees H. Taal*[4]

[1] Signal and Information Processing Lab, Delft University of Technology, The Netherlands
[2] Oticon A/S, Denmark
[3] Department of Electronic Systems, Aalborg University, Denmark
[4] Applied Sensor Technologies, Philips Research, The Netherlands

ABSTRACT

While most contributions on speech reinforcement only consider the presence of environmental noise, late reverberation can also severely degrade the intelligibility of speech. In this paper we address the problem of speech reinforcement in noisy and reverberant environments. We use a short-time version of a recently presented approximation of the speech intelligibility index, which we optimize locally. The resulting time-frequency dependent amplification depends on both the noise and late reverberation power spectral density. The latter is estimated using the Polack model and assumes that prior knowledge of the room geometry is available. Speech intelligibility improvements of around 20% are observed.

Index Terms— Speech reinforcement, speech intelligibility, additive noise, late reverberation, approximated SII

1. INTRODUCTION

Users of speech communication systems like public address or conference systems often experience a degradation of speech intelligibility due to the presence of environmental noise in the vicinity. Pre-processing the speech signal prior to playback in the noisy environment can partly restore the intelligibility, even in challenging noisy conditions. This problem attracted an increased research interest over the last few years, *e.g.*, [1, 2, 3, 4, 5, 6, 7, 8]. Among recent contributions we see a growing interest to optimize for quantitative measures of speech intelligibility, *e.g.*, [3, 4, 5, 6, 7, 9]. An often used quantitative measure is the speech intelligibility index (SII) [10], for example [3, 4, 7]. The SII can predict the effects of additive stationary noise on intelligibility by comparing the long-term average speech and noise energy within critical bands. Extensions to model speech intelligibility in non-stationary noise have also been proposed, *e.g.*, based on a short-time variant of SII, often referred to as extended SII (ESII) [11].

Besides SII-based measures, other measures have been considered as well. For example, in [9] a speech intelligibility model based on the mutual information between the original and the processed noisy speech was presented and optimized, while in [6] a spectro-temporal perceptual distortion measure presented in [12] was optimized in order to optimally redistribute speech energy over frequency and time.

This work was supported in part by the Dutch Technology Foundation STW and Bosch Security Systems B.V.

Besides background noise, late reverberation can also severely degrade the speech intelligibility [13], in particular in applications like public address systems. Despite this fact, the case with both additive noise and reverberation has only rarely been treated. Two examples where only reverberation is taken into account are [14, 15].

In this paper we investigate whether late reverberation and additive noise can be jointly taken into account, while optimizing for a quantitative measure of intelligibility under an energy constraint. The energy constraint is used to overcome too loud sounds that might damage the human auditory system or loudspeakers. As optimization criterion we use an SII-based measure. Despite the simplicity of the original SII model, constrained optimization of SII leads to a non-convex optimization problem, which is a reason to introduce approximations to the original SII model. We employ the approximation of the SII model presented in [7], which we refer to as the approximated SII (ASII). Because late reverberation is non-stationary, we use a short-time version of ASII denoted by ASII_{ST}. As global optimization would result in a non-convex problem, we perform the optimization locally per time frame for all frequency bands taking the reverberation generated by a fixed number of past time frames into account.

2. NOTATION AND ASSUMPTIONS

Let $x(n)$ denote the observed noisy reverberated speech process with time-sample index n, given by

$$x(n) = (h * s_p)(n) + w(n) = e(n) + z(n) + w(n), \quad (1)$$

with h the time-invariant room impulse response, $s_p(n)$ a processed version of the original speech $s(n)$, $w(n)$ additive noise uncorrelated with $s(n)$, and, $e(n)$ and $z(n)$ the early and late reverberation of the processed speech, respectively. Processing will be done per critical band and time frame. Let g_i be the impulse response of the ith critical band filter with its discrete Fourier transform (DFT) for frequency bin k given by $G_i(k)$. Further, we denote the DFT coefficient of a speech frame at the loudspeaker for frequency-bin k and time frame starting at sampling-index m by $S(m,k)$. Similarly we define the noise and late reverberant DFT coefficients by $W(m,k)$ and $Z(m,k)$, respectively. The speech, noise, and late reverberant energy per critical band and time frame are then given by $\mathcal{S}^2(m,i) = \sum_k |S(m,k)|^2|G_i(k)|^2$, $\mathcal{W}^2(m,i) = \sum_k |W(m,k)|^2|G_i(k)|^2$

and $\mathcal{Z}^2(m,i) = \sum_k |Z(m,k)|^2 |G_i(k)|^2$, respectively. The critical band energy of the processed speech is given by $\alpha^2(m,i)\mathcal{S}^2(m,i)$, with $\alpha(m,i)$ the time frame and critical band dependent amplification. Let $E[\cdot]$ be the statistical expectation operator. DFT domain variances of the speech, noise and late reverberation are then defined as $\sigma_S^2(m,k) = E[|S(m,k)|^2]$, $\sigma_W^2(m,k) = E[|W(m,k)|^2]$ and $\sigma_Z^2(m,k) = E[|Z(m,k)|^2]$, respectively, and the corresponding variances per critical band and time frame are given by $\sigma_{\mathcal{S}}^2(m,i) = \sum_k \sigma_S^2(m,k)|G_i(k)|^2$, $\sigma_{\mathcal{W}}^2(m,i) = \sum_k \sigma_W^2(m,k)|G_i(k)|^2$, and $\sigma_{\mathcal{Z}}^2(m,i) = \sum_k \sigma_Z^2(m,k)|G_i(k)|^2$ respectively.

We will use the ASII_{ST} to find the locally optimal amplification factors $\alpha(m,i)$ under an energy constraint. The ASII was proposed in [7] because constrained optimization of the SII constitutes a non-convex problem. Whereas the original ASII models the speech intelligibility as a function of the long-term signal-to-noise ratio (SNR) per critical band, we use a short-time variant of the ASII, (similar to ESII [11]), to take the time time-varying nature of late reverberation into account. We therefore assume speech and noise processes to be stationary and ergodic within a time-frequency unit and use SNR estimates per time frame and critical band.

Using $\sigma_{\mathcal{S}}^2(m,i)$ and $\sigma_{\mathcal{W}}^2(m,i)$, the SNR is given by

$$\xi(m,i) = \frac{\sigma_{\mathcal{S}}^2(m,i)}{\sigma_{\mathcal{W}}^2(m,i)}. \tag{2}$$

Let γ_i denote the critical-band importance function given in [10]. In analogy to the ASII presented in [7], the ASII_{ST} for a time-frame m is then given by

$$\text{ASII}_{\text{ST},m} = \sum_i \gamma_i \frac{\xi(m,i)}{\xi(m,i)+1}. \tag{3}$$

3. ASII INCLUDING LATE REVERBERATION

Although early reflections contribute to speech intelligibility (e.g., [16]), we consider for simplicity a worst case scenario and neglect these. We set the direct part of the processed signal as the desired signal and assume that the damping Δ from loudspeaker to listener location is inversely proportional to the distance. Further we define $n_\Delta = \Delta^{-1} f_s c^{-1}$ as the delay from loudspeaker to listener location in samples with f_s the sampling frequency and c the speed of sound. Let $n_0 = n_\Delta + \tau$ denote the starting sample of the late reflections of the impulse response, with τ a 50 ms pause between the direct path and the start of the late reverberation [16].

Let $\sigma_{\mathcal{Z}_p}^2(m,i)$ denote the late reverberation of the processed speech per critical band and time frame. To take both late reverberation and additive noise into account, we define the SNR as the ratio of the variance of the processed speech per critical band and time frame (i.e., after amplification by $\alpha(m,i)$) at the listener location, that is, $\alpha^2(m,i)\sigma_{\mathcal{S}}^2(m,i)\Delta^2$, and the sum of the noise variance and late reverberation variance present n_0 samples after play out by the loudspeaker. That is,

$$\xi(m+n_\Delta,i) = \frac{\alpha^2(m,i)\sigma_{\mathcal{S}}^2(m,i)\Delta^2}{\sigma_{\mathcal{Z}_p}^2(m+n_0,i) + \sigma_{\mathcal{W}}^2(m+n_\Delta,i)}. \tag{4}$$

Next to $\sigma_{\mathcal{W}}^2$, we also need the late reverberation variance $\sigma_{\mathcal{Z}_p}^2$, which will be derived below.

To model the late reflections of the impulse response we use the Polack model [17], that is,

$$h(l) = a^{l-n_0} u(l-n_0),\ l \geq n_0, \tag{5}$$

with $u(l)$ an uncorrelated white stationary Gaussian noise process with variance σ_u^2 and a a damping factor.

Let $v(\cdot)$ be a length-N window. Using a result from [18], a useful expression for the late reverberation DFT coefficient can be derived under the assumption that the impulse response is time-invariant during a time-frame [18]. That is,

$$\begin{aligned} Z(m,k) &= \sum_{n=m}^{m+N-1} v(n-m) \sum_{l=n_0}^{+\infty} h(l)s(n-l)e^{-j2\pi k\frac{(n-m)}{N}} \\ &\approx \sum_{l=n_0}^{+\infty} h(l) \sum_{n=m}^{m+N-1} v(n-m)s(n-l)e^{-j2\pi k\frac{(n-m)}{N}} \\ &= \sum_{l=n_0}^{+\infty} h(l)S(m-l,k). \end{aligned}$$

Let $R = N/2$ be the frame shift. Using the assumption that speech is stationary over a time frame, and that u and S are two independent processes, we obtain the following expression for $\sigma_Z^2(m,k)$ using the geometric series

$$\sigma_Z^2(m,k) = \sigma_u^2 \sum_{p=0}^{+\infty} \sigma_S^2(m-n_0-pR,k) \frac{a^{2pR}(1-a^{2N})}{1-a^2}. \tag{6}$$

Using the diffuse room impulse response energy $\rho^2 = \sum_{l=n_0}^{+\infty} E[h^2(l)]$, σ_u^2 is given by $\sigma_u^2 = (1-a^2)\rho^2$ [19]. Taking the amplification $\alpha^2(m,i)$ into account we finally obtain

$$\begin{aligned} \sigma_{\mathcal{Z}_p}^2(m,i) &= \rho^2\left(1-a^{2N}\right) \\ &\times \sum_{p=0}^{P-1} a^{2pR}\alpha^2(m-n_0-pR,i)\sigma_{\mathcal{S}}^2(m-n_0-pR,i), \end{aligned} \tag{7}$$

where, compared to (6), we have truncated the summation over p to P time frames (reflecting the late reverberation).

4. OPTIMIZING $\text{ASII}_{\text{ST},M}$

In this section we maximize $ASII_{ST}$ locally for all critical bands in time-frame m, by finding optimal values of $\alpha^2(\ell,i)$ for all critical bands i and $\ell \in \mathcal{L}$ with $\mathcal{L} = \{m-(P-1)R, m-(P-2)R, ..., m\}$. For time-frame m we then get the cost function

$$J = \sum_i \frac{\gamma_i \alpha^2(m,i)\sigma_{\mathcal{S}}^2(m,i)\Delta^2}{\alpha^2(m,i)\sigma_{\mathcal{S}}^2(m,i)\Delta^2 + \sigma_{\mathcal{W}}^2(m+n_\Delta,i) + \sigma_{\mathcal{Z}_p}^2(m+n_0,i)},$$

with $\sigma_{\mathcal{Z}_p}^2(m,i)$ as in (7). Notice that $\sigma_{\mathcal{Z}_p}^2(m+n_0,i)$ depends on $\alpha^2(\ell,i)$ with $\ell \in \mathcal{L}$.

For each critical band i, both the numerator and denominator of the terms in J consist of positive terms only. Setting any of the $\alpha^2(\ell,i)$ with $\ell \in \mathcal{L} \setminus \ell = m$ to zero (i.e., any α^2 except $\alpha^2(m,i)$)) will always increase the value of J. Setting all $\alpha^2(\ell,i)$ with $\ell \in \mathcal{L} \setminus \ell = m$ to zero, we obtain the maximum of J with respect to $\alpha^2(\ell,i)$ for $\ell \in \mathcal{L} \setminus \ell = m$, leaving only

$\alpha^2(m,i)$ undetermined. This argumentation follows from the fact that the cost function has only has a local view on the problem and does not take distortions into account that result from setting gains to zero in past time frames. With this result we can define a new (simplified) cost function as

$$J_1 = \sum_i \frac{\gamma_i \alpha^2(m,i)\sigma_S^2(m,i)\Delta^2}{\alpha^2(m,i)\sigma_S^2(m,i)\left(\Delta^2+\rho^2\left(1-a^{2N}\right)\right)+\sigma_{\mathcal{W}}^2(m+n_\Delta,i)}.$$

As $\max J \geq J_1 \geq J$, we come to the following problem formulation

$$\max_{\alpha^2(m,i)\,\forall\, i} J_1 \tag{8}$$

$$s.t. \sum_i \sum_{\forall \ell \in \mathcal{L}} \alpha^2(\ell,i)\sigma_S^2(\ell,i) = \sum_i \sum_{\forall \ell \in \mathcal{L}} \sigma_S^2(\ell,i) \tag{9}$$

$$\alpha^2(\ell,i) = 0, \text{ for } \ell \in \mathcal{L} \setminus \ell = m \text{ and} \tag{10}$$

$$\alpha^2(m,i) \geq 0, . \tag{11}$$

Eqs. (8)-(11) form a convex optimization problem as the objective function is concave in $\alpha^2(m,i)$ and the constraints are all linear in $\alpha^2(\ell,i)$ with $\ell \in \mathcal{L}$. The Karush-Kuhn-Tucker (KKT) conditions are therefore necessary and sufficient conditions to find a maximum. Calculating the KKT conditions and solving for $\alpha^2(m,i)$ we finally find

$$\alpha^2(m,i) = \frac{\max\left(\frac{\sigma_{\mathcal{W}}(m+n_\Delta,i)\sqrt{\gamma_i}\Delta}{\sqrt{\nu}} - \sigma_{\mathcal{W}}^2(m+n_\Delta,i), 0\right)}{\sigma_S^2(m,i)(\Delta^2+\rho^2\left(1-a^{2N}\right))} \tag{12}$$

$$\alpha^2(\ell,i) = 0, \text{ for } \ell \in \mathcal{L} \setminus \ell = m \tag{13}$$

$$\sum_i \max\left(\frac{\sigma_{\mathcal{W}}(m+n_\Delta,i)\sqrt{\gamma_i}\Delta}{\sqrt{\nu}} - \sigma_{\mathcal{W}}^2(m+n_\Delta,i), 0\right) = r(\Delta^2+\rho^2\left(1-a^{2N}\right)). \tag{14}$$

Calculating $\alpha^2(m,i)$ in (12) depends on finding the value for ν using a root finding procedure on (14). Among other methods, this can be done using a bisection method, followed by substitution of ν in (12). Also, note that (8)–(11) generalizes the problem statement in [7]. For $\rho = 0$ and $P = 1$ (i.e., no reverberation), the solution is identical to the one in [7].

Similar to [7], the proposed algorithm only amplifies critical bands that are relevant for intelligibility. If the SNR and amplification within the energy constraint will not help to increase the objective function, these bands will automatically be clipped to zero. Compared to [7], two differences are present. At first, if late reverberation is present ($\rho > 0$), $\alpha^2(m,i)$ is in all bands decreased by $(\Delta^2+\rho^2\left(1-a^{2N}\right))$, taking into account that amplifying speech will automatically increase the distortion introduced by the late reverberation. The resulting energy that is left is used to amplify frequency bands that are otherwise being clipped to zero.

Secondly, parameter P can be used to introduce dynamic gain compression. For noisy speech, dynamic gain compression is known to increase the intelligibility [20, 8]. For reverberant speech, dynamic gain compression works as a steady state suppressor where for increasing P, energy of low energy transients is increased over the stationary high energy regions. As described above, per time-frequency unit, P gains are available. $P-1$ of these gains equal zero. Taking the average to combine the P gains per time-frequency unit, the processed speech energy per time frame becomes $\sum_{\forall i} \sigma_{\mathcal{S}_p}^2(m,i) = \frac{r}{P}$. Setting $r = \sum_{\forall l \in \mathcal{L}} \sum_{\forall i} \sigma_S^2(l,i)$, implies the frame energy to be set to the average energy over the last P time-frames.

5. EXPERIMENTAL RESULTS

In this section we present a comparison based on instrumental measures as well as a speech intelligibility listening test, evaluating the presented algorithm and reference methods. As reference methods we use the steady state suppressor of Hodoshima et al. [21] referred to as *Hodo06* and the method presented in [7] referred to as *Taal13* as this is a special case of the presented algorithm that does not take late reverberation into account. The output of all algorithms is guaranteed to equal the input. For the proposed approach and *Taal13* we use exponential smoothing with $\beta = 0.996$ to measure the speech variance $\sigma_S^2(m,i)$. These two algorithms also depend on the noise power spectral density (PSD). To eliminate estimation errors on the noise PSD, we measure the noise PSD based on the noise-only signal using an ideal voice activity detector. Furthermore, we assume the room dimensions and $T60$ reverberation time to be known. Based on this we can compute the diffuse room impulse response energy ρ. The room volume is set to approximately $10 \times 28 \times 4\,m^3$ ($L \times W \times H$). To calculate ρ, we make use of the direct-to-reverberation ratio [22], in combination with Sabine's equation (see e.g. [22]). Based on initial experiments, the value for P is set to $P = 3$, introducing dynamic range compression.

The speech level is calibrated at 62.35 dB SPL, with 120 dB SPL as the maximum playback level. The noisy reverberant signal $x(n)$ is generated according to (1), where the convolution is performed in the DFT domain. We neglect early reflections in modeling the room impulse response and model the direct path by a delta impulse with height Δ. The late reflections are generated using the Polack model [17], where the exponential decay is given by $a = 10^{-\frac{3}{T_{60} f_s}}$. In all experiments, the distance between loudspeaker and listener is is set to $\Delta = 1/d$ with $d = 5$. The proposed approach is used on a frame-by-frame basis with 32-ms frames taken with 50% overlap and windowed with a square-root Hann window. All signals are sampled at 16 kHz.

5.1. Instrumental Comparison to Reference Methods

In generating the instrumental comparison, more than 5 minutes of speech was used, originating from the Timit database [23]. The speech signals are concatenated and degraded by stationary speech-shaped noise and babble noise at SNRs of -5 dB and 0 dB, measured between the original unprocessed speech (at the loudspeaker) and the background noise.

To measure the intelligibility improvements we use two instrumental measures. We use ASII_{ST} as this is the measure that is being optimized in this paper. In addition we use the extended or short-time SII, denoted as ESII [11] as it is known to be a good predictor of intelligibility under time-varying (uncorrelated) noise maskers. Moreover, this is the measure that is essentially approximated by ASII_{ST}.

The experimental results are depicted in Figs. 1-2 in terms of instrumental intelligibility improvement over the unprocessed signal as a function of the T60 ranging from 0 seconds

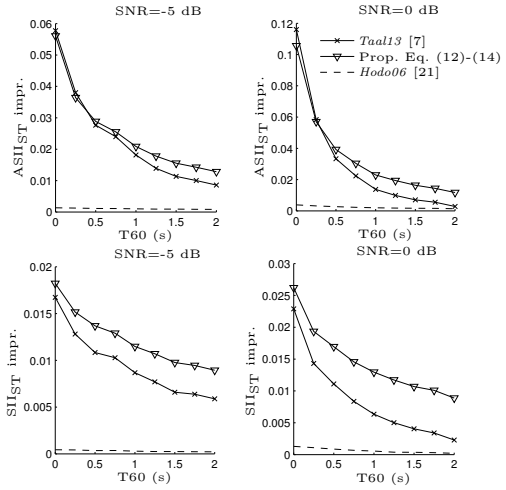

Fig. 1. Instrumental intelligibility in terms of ESII and ASII_{ST} improvement for speech degraded by speech-shaped noise.

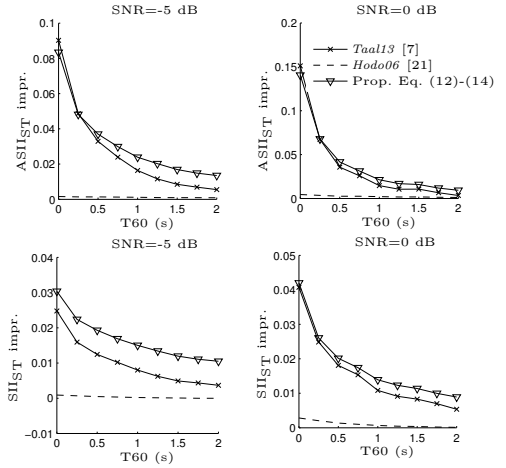

Fig. 2. Instrumental intelligibility in terms of ESII and ASII_{ST} improvement for speech degraded by babble noise.

to 2 seconds. The proposed method improves performance compared to *Taal13* over the whole range of T60 values, both SNRs and noise types. When there is no reverberation ($T60 = 0$), the proposed method and the method from [7] differ only due to the dynamic range compression that is performed in the proposed method with $P = 3$. With increasing T60, the improvement of the proposed method over *Taal13* increases slightly as a function of T60. According to the used instrumental intelligibility measures, *Hodo06* shows only minor improvements over the unprocessed signal.

5.2. Intelligibility Listening Test

In this section we present speech intelligibility listening test results of *Taal13*, the proposed approach and the unprocessed signal. We adopted the Dutch closed speech-in-noise intelligibility test [24], which we used under noisy reverberant conditions. The test material consists of five-word sentences with correct grammatical structure. The listener selects via a graphical interface the words that were understood. The possible words are arranged in a 10-by-5 matrix on a computer screen, such that the ith column contains exactly the 10 possible alternatives for the ith word. The T60 time of the room impulse response was set to 1 second. As noise source we used speech-shaped noise at SNRs of -2, 0, 2 and 4 dB with respect to the original signal as present at the loudspeaker.

Seven native Dutch speaking subjects participated in the test. The order of presenting the different algorithms and the SNRs were randomized, with each sentence being used only once. For each test person, all processing conditions were repeated four times. The signals were presented diotically through head-phones (Sennheiser HD 600).

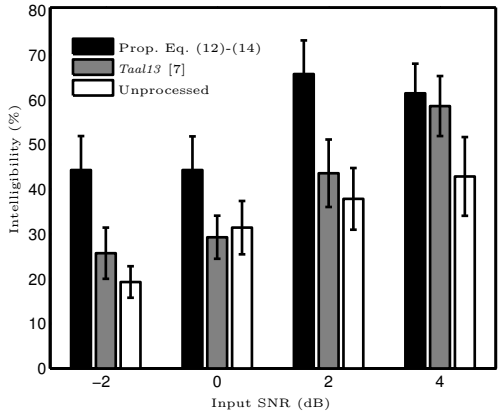

Fig. 3. Intelligibility listening test results.

The average intelligibility scores with standard error of the mean are shown in Fig. 3. This shows that under all conditions, the proposed method improves over *Taal13* and the unprocessed signals. To determine the statistical significance, a t-test [25] with a significance level $\alpha = 0.05$ was performed. From this t-test it follows that the proposed method is always significantly better intelligible than the unprocessed signals. Compared to *Taal13*, the proposed method is significantly better for all SNRs, except at the SNR of 4 dB.

6. CONCLUSIONS

In this paper we presented an algorithm for speech intelligibility enhancement in noisy reverberant conditions. We employed a short-time version of a recently presented approximation of the SII, facilitating constrained optimization. For mathematical tractability, we optimize the ASII locally, taking a segment of past time frames into account. The late reverberation is modeled using the Polack model.

The optimization results in critical-band and time-frame depending amplification factors that redistribute the energy across frequency taking into account the PSDs of the noise and the generated late reverberation. Instrumental experiments and intelligibility listening tests show an increase of the proposed approach over the unprocessed condition, as well as over neglecting the presence of late reverberation. An increase of the intelligibility of approximately 20% is observed.

7. REFERENCES

[1] C. Tantibundhit, J. R. Boston, C. C. Li, J. D. Durrant, S. Shaiman, K. Kovacyk, and A. El-Jaroudi, "New signal decomposition method based speech enhancement," *Signal Processing*, vol. 87, pp. 2607 – 2628, 2007.

[2] B. Sauert and P. Vary, "Near end listening enhancement: speech intelligibility improvement in noisy environments," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2006, vol. 1, pp. 493–496.

[3] B. Sauert and P. Vary, "Near end listening enhancement optimized with respect to speech intelligibility index," in *EURASIP Europ. Signal Process. Conf. (EUSIPCO)*, 2009, vol. 17, pp. 1844–1848.

[4] B. Sauert and P. Vary, "Near end listening enhancement optimized with respect to speech intelligibility index and audio power limitations," in *EURASIP Europ. Signal Process. Conf. (EUSIPCO)*, 2010, pp. 1919–1923.

[5] B. Sauert and P. Vary, "Recursive closed-form optimization of spectral audio power allocation for near end listening enhancement," in *ITG-Fachtagung Sprachkommun.* 2010, VDE VERLAG GmbH.

[6] C. H. Taal, R. C. Hendriks, and R. Heusdens, "A speech preprocessing strategy for intelligibility improvement in noise based on a perceptual distortion measure," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*. IEEE, 2012, pp. 4061–4064.

[7] C. H. Taal, J. Jensen, and A. Leijon, "On optimal linear filtering of speech for near-end listening enhancement," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 225 – 228, 2013.

[8] M. Cooke, C. Mayoe, and C. Valentini-Botinhao, "Intelligibility enhancing speech modifications: the hurricane challenge," *ISCA Interspeech*, 2013.

[9] W. B. Kleijn and R. C. Hendriks, "A simple model of speech communication and its application to intelligibility enhancement," *IEEE Signal Process. Lett.*, Online available Sept. 2014.

[10] The American National Standards Institute, *American National Standard Methods for the Calculation of the Speech Intelligibility index*, Ansi S3.5-1997 edition.

[11] K. S. Rhebergen and N. J. Versfeld, "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," *J. Acoust. Soc. Amer.*, vol. 117, no. 4, pp. 2181–2192, April 2005.

[12] C. H. Taal, R. C. Hendriks, and R. Heusdens, "A low-complexity spectro-temporal distortion measure for audio processing applications," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 5, pp. 1553–1564, 2012.

[13] A. C. Neuman, M. Wroblewski, J. Hajicek, and A. Rubinstein, "Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults," *Ear and Hearing*, vol. 31, no. 3, pp. 336–344, 2010.

[14] A. Kusumoto, T. Arai, K. Kinoshita, N. Hodoshima, and N. Vaughan, "Modulation enhancement of speech by a pre-processing algorithm for improving intelligibility in reverberant environments," *ELSEVIER Speech Commun.*, vol. 45, no. 2, pp. 101–113, 2005.

[15] N. Hodoshima, T. Arai, A. Kusumoto, and K. Kinoshita, "Improving syllable identification by a preprocessing method reducing overlap-masking in reverberant environments," *J. Acoust. Soc. Amer.*, vol. 119, pp. 4055, 2006.

[16] J. S. Bradley, H. Sato, and M. Picard, "On the importance of early reflections for speech in rooms," *J. Acoust. Soc. Amer.*, vol. 113, no. 6, pp. 3233–3244, 2003.

[17] J. D. Polack, *La transmission de l'energie sonore dans les salles*, Ph.D. thesis, Universitè du Maine, Le Mans, France, 1988.

[18] J. S. Erkelens and R. Heusdens, "Correlation-based and model-based blind single-channel late-reverberation suppression in noisy time-varying acoustical environments," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 7, pp. 1746–1765, 2010.

[19] J. Crespo and R. C. Hendriks, "Multizone speech reinforcement," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 1, 2014.

[20] 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, Signal Process.*, vol. 24, no. 4, pp. 277–282, 1976.

[21] N. Hodoshima, T. Arai, A. Kusumot, and K. Kinoshita, "Improving syllable identification by a preprocessing method reducing overlap-masking in reverberant environments)," *J. Acoust. Soc. Amer.*, vol. 119, no. 6, pp. 4055–4064, 2006.

[22] M. J. Crocker, *Handbook of noise and vibration control*, John Wiley & Sons, 2007.

[23] J. S. Garofolo, "DARPA TIMIT acoustic-phonetic speech database," *National Institute of Standards and Technology (NIST)*, 1988.

[24] R. Houben, J. Koopman, H. Luts, K. C. Wagener, A. van Wieringen, H. Verschuure, and W. A. Dreschler, "Development of a dutch matrix sentence test to assess speech intelligibility in noise," *International Journal of Audiology, Early Online*, vol. 127, pp. 1–4, 2014.

[25] D. J. Sheskin, *Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, 3rd edition edition, 2004.