Adaptive Gain Methods to Improve Speech Intelligibility under Reverberation

F. de S. Farias and R. Coelho

Abstract— **The reflection of an acoustic signal on walls or objects in an enclosed environment is perceived as reverberation. It is present daily in conference rooms, tunnels and any closed spaces. The presence of reverberation can degrade speech intelligibility and the performance of tasks that depend on it, such as speech or speaker recognition. Many methods were proposed to attenuate this degradation. A family of such methods acts on clean speech, applying gains on parts of the signal in order to improve intelligible when reverberated. These methods are called *adaptive gain methods*. This study aims to evaluate the effect of two adaptive gain methods, Adaptive Gain Control (AGC) and Steady-State Suppression (SSS), in speech intelligibility. The evaluation uses four objective measures: Coherence Signal Intelligibility Index (CSII), Short Time Objective Intelligibility (STOI), Speech Reverberation to Modulation Ratio (SRMR) and Weighted Short Time Objective Intelligibility (WSTOI). Results show that AGC improves speech intelligibility in studied conditions, while SSS degrades it, a result in line with subjective measures found in the literature.**

Keywords— **Intelligibility, Reverberated Speech, Reverberation**

I. Introduction

The reflections of an acoustic signal on walls and objects in the environment are perceived as reverberation. The first and stronger reflections are called *early reflections* (ER), while the numerous and attenuated reflections that come later are called *late reflections* (LR). Reverberation is often observed in real life situations, mostly in closed spaces such as in subway stations, empty rooms or caves.

Studies of speech perception in rooms show that reverberation degrades speech intelligibility in two forms: self-masking, which is the temporal blurring of the signal within each phoneme, and overlap-masking, which is the masking of phonemes by the reflections from previous phonemes [1], [2]. This degradation impacts on tasks such as automatic speaker verification and speech recognition [3], [4].

Many methods were proposed to mitigate the degradation of reverberation on speech intelligibility. One of the first and still most popular class of methods is the *inverse filtering*, which is the passing of a reverberated signal through a filter that inverts the effects of reverberation [5], [6]. However, this method relies on the inversion of the Room Impulse Response (RIR), the representation of the reaction of a room to an impulsive sound. This task is computationally expensive, especially in highly reverberant conditions.

Another class of methods, the *adaptive gain* methods address this problem by applying a gain on the speech signal before it is reverberated. The modification is optimized based on knowledge about the room and the speech signal, so the processed speech is more intelligible than the original when reverberated [7], [8]. However, evaluation and comparison of such methods has been done only in one highly reverberant condition [9].

The goal of this work is to study the effect of two adaptive gain methods, Steady-State Suppression (SSS) [10] and Adaptive Gain Control (AGC) [9] on speech intelligibility under several previously unexplored reverberation conditions, thus adding to the understanding of said methods. The evaluation is conducted on a subset of the TIMIT database, on two different rooms and seven different reverberation times (T_{60}), from 0.6s to 1.8s. The intelligibility is measured using four objective metrics: Short-Time Objective Intelligibility (STOI) [11], Weighted Short-Time Objective Intelligibility (WSTOI) [12], Speech Reverberation Modulation Ratio (SRMR) [13] and Coherence Speech Intelligibility Index (CSII) [14]. Results show that AGC improves speech intelligibility on T_{60} longer than 1.4s. Of the two rooms studied, it shows a significant improvement in the biggest. SSS decreases intelligibility for all studied conditions, a result that reflects previous experiments [7], [10].

The remainder of this letter is organized as follows. Section II describes two adaptive gain methods evaluated in this work. Section III summarizes the objective intelligibility measures used in the evaluation. In Section IV, the experimental results are presented, followed by conclusion in Section V.

II. Adaptive Gain Methods

This section briefly describes the adaptive gain methods that are evaluated in this work.

A. Steady-State Suppression (SSS)

SSS [10] is an adaptive gain method that focus on suppressing the steady part of the reverberant signal. It was developed to reduce the intelligibility degradation on reverberated speech.

This method suppresses spectrum regions deemed less important to intelligibility based on a parameter D [15]. This parameter measures the spectral transition of the signal. When it is lower than a certain threshold, the signal is attenuated.

The calculation of D follows these steps:

1) *Extraction of the temporal envelope:* It is performed first splitting the signal in $\frac{1}{3}$ octave bands, followed by using the Hilbert transform in each band.
2) *Calculation of regression coefficients:* First, the signal is downsampled. Then, five adjacent values of the logarithmic time trajectory are used to calculate the regression coefficients.
3) *Calculation of parameter D:* the parameter is obtained by mean square of the regression coefficients.

Felipe de S. Farias, Rosângela F. Coelho Programa de Pós Graduação em Engenharia de Defesa, Instituto Militar de Engenharia (IME), Rio de Janeiro-RJ, Brazil, E-mails: felipe.farias@ime.eb.br,coelho@ime.eb.br. This work was partially supported by CNPq (307866/2015-7).

$$D = \frac{1}{N} \sum |c_i|^2 \tag{1}$$

where N is the number of bands and c_i is the regression coefficient in band i.

Experiments with consonant recognition in reverberant conditions show that the method improves intelligibility in low reverberation for hearing impaired subjects, but degrades intelligibility for normal hearing subjects [7], [10].

B. Adaptive Gain Control (AGC)

AGC [9] is a method that optimizes the signal gain using a nonstationarity measure. It aims to reduce the overlap-masking in nonstationary regions of the reverberant speech [16].

In this method, the signal gain is designed to optimize a speech-to-late-reverberation ratio (SLRR) by Lagrange multiplier.

$$y(x) = c_1 x + c_2 x^b + \frac{l}{2b}(l\lambda - 2b) \tag{2}$$

where λ is the Lagrange multiplier and c_1 and c_2 are determined by the boundary conditions.

The calculation of the gain follows these steps:

1) *Nonstationarity estimation:* The nonstationarity index used in this work is the normalized distance between the mel-frequency cepstral coefficients (MFCCs) in adjacent frames.

$$\xi_i = \frac{||m_i - m_{i-1}||}{||m_i|| + ||m_{i-1}||} \tag{3}$$

2) *Computation of LR power:* The LR signal is estimated by convolving the original signal with a pulse-train model for the LR part of the room impulse response (RIR) based on the velvet noise [17].
3) *Computation of the gain:* This is done in three parts. The first is the computation of the logarithm of ν_ξ, a sigmoid function of the nonstationarity index ξ.

$$log(\nu_\xi) = q(\xi|s, log(\beta), log(\alpha)) \tag{4}$$

where s is the slope factor of the function, α and β determine the range of interest.

The gain penalty λ_{ν_ξ} is calculated from each ν_ξ through the equation:

$$\lambda_\xi = \frac{2b * (\rho - 1)(\alpha^b \nu - \alpha \nu^b)}{l^2(\nu^b - \alpha_b - b(\nu - \alpha)\psi^{b-1})} + \frac{2b}{l} \tag{5}$$

where l is the power of the LR part of the signal and b is the shape parameter in the probability density function of x.

The second step is the selection of λ, where two cases apply depending on the value of l. For $l \leq \tilde{\lambda}$:

$$\lambda = max(\tilde{\lambda}, \lambda_\xi) \tag{6}$$

where $\tilde{\lambda}$ is the critical value in which $y = \beta$ and $\lambda = \tilde{\lambda}$. For $l > \tilde{l}$,

$$\lambda = \lambda_{\overline{\nu}}, \left(log(\overline{\nu} = q(\frac{\lambda_{\nu_\xi}}{\tilde{\lambda}}|s, log(\nu_\xi), log(\alpha)) \right) \tag{7}$$

constrain the maximum boosting power to $\overline{\nu} \in (\alpha, \nu_\epsilon)$.

From that, $y_i(x)$ can be obtained through equation (2) and the gain is then defined by

$$g_i = \sqrt{\frac{y_i}{x_i}} \tag{8}$$

4) *Smoothing of the gain:* The smoothing is made using two sigmoid functions u and d which serve, respectively, as the upper and lower bounds of the gain, effectively reining the amount of gain modification the algorithm does.

$$g_i = \begin{cases} min(u(\xi_i, g_i), g_i), & g_i > 1 \\ min(d(\xi_i), g_i), & g_i \leq 1 \end{cases} \tag{9}$$

Subjective evaluations compare this method with SSS in one room, under one strong reverberant condition. Results show show that speech processed with AGC is more intelligible than unprocessed reverberated speech or speech processed with SSS.

III. Objective Intelligibility Measures

This section describes objective intelligibility measures (OIM), STOI [11], WSTOI [12], SRMR [13] and CSII [14], used to evaluate the reverberated and processed speech.

A. Short Time Objective Intelligibility (STOI)

STOI was developed to predict the intelligibility of noisy speech processed by time-frequency weighting masks, such as speech enhancement methods. Experiments presented in [11] show it yields high correlation with the intelligibility of speech distorted by additive noise and processed by different types of speech enhancement algorithms.

The STOI score is the correlation coefficient between the spectral envelopes of clean and relative enhanced signal. This measure is obtained through six steps:

1) *Segmentation:* the clean and the distorted signals are separated in short time Hann-windowed frames. Spectral decomposition of both signals using the FFT.
2) *Extraction of silence:* silent regions of the clean speech are excluded from the computation. This selection is based on the energy of each frame compared with the frame with highest energy of the signal.
3) *Grouping:* the signals are grouped in a vector of 30 frames across 15 $1/3$-octave bands.
4) *Normalization:* the distorted signal is normalized in comparison to the clean signal.
5) *Calculation:* STOI is calculated as an average of the correlations between each group in the clean and distorted signal.
6) *Mapping:* to translate the values of STOI to word recognition ratio (WRR), a mapping function must be used. It takes the form of

$$STOImap = \frac{100}{1 + exp(a * STOI + b)} \tag{10}$$

in which $STOImap$ is the predicted intelligibility score, $STOI$ is the output of the algorithm, a and b are two free parameters, adjusted according to the experiment.

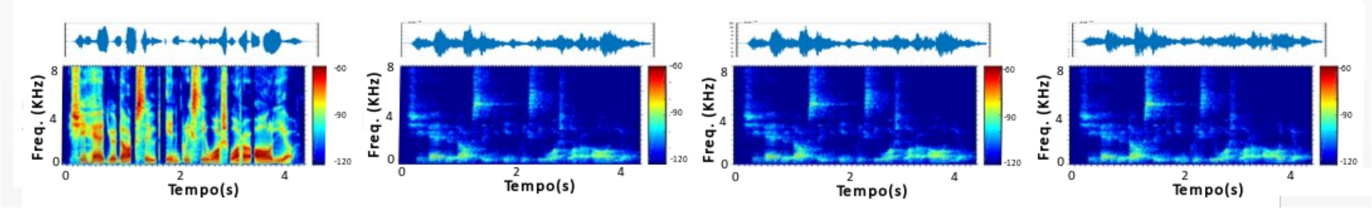

Fig. 1: Waveforms and spectrograms of the following signals, respectively: clean speech, reverberated speech (Petkov room, $T_{60} = 1.8$s), reverberated speech processed using SSS and reverberated speech processed using AGC.

B. Weighted Short Time Objective Intelligibility (WSTOI)

WSTOI uses mutual information to provide an estimate of the intelligible content of the speech signal and give a better prediction of it's intelligibility [12]. It is a measure built on top of STOI, shown in Section III-A. The main difference from STOI is the use of a mutual information estimation to weight the contribution of each time-frequency cell and improve the intelligibility prediction. As it is shown in [12], this measure obtained better accuracy than STOI in speech distorted by several types of noise in several SNRs.

This metric follows the same steps as STOI, adding the following processing in the clear speech, after the *Segmentation* step:

1) *Prediction:* the next phoneme is predicted using a LPC predictor that uses the previous two phonemes.
2) *Mutual information calculation:* the mutual information between the clear and predicted speech is calculated.

The mutual information is used to weight the correlation between each group of clear and distorted signals, giving then a different and more accurate score. One side effect of the weighting is that the silence extraction step is no longer necessary.

C. Speech Reverberation Modulation Energy Ratio (SRMR)

SRMR is an adaptive non intrusive measure of quality of reverberated and processed speech signals. It is the ratio between the average of the modulation energy of the first four modulation bands of a signal and the average of rest of the bands, as can be seen in equation (11). Studies show this ratio has great correlation with perceptual quality and intelligibility measures in reverberant conditions [13].

To calculate this ratio, the following steps are required:

1) *Separation:* the signal is separated in 23 bands, resembling the function of the human cochlea.
2) *Extraction of envelope:* for each frequency band, the temporal envelope is extracted and split into frames.
3) *Calculation of Modulation Energy:* the modulation spectral energy is calculated as the squared magnitude of the discrete Fourier transform of the the temporal envelope.
4) *Calculation of the Modulation Ratio:* the modulation frequency bins are grouped in 8 bands. $\overline{\xi_k}$ is the average of the modulation energy in modulation band k. The ratio is then given by:

$$SRMR = \frac{\sum_{k=1}^{4} \overline{\xi_k}}{\sum_{k=5}^{K} \overline{\xi_k}} \tag{11}$$

D. Coherence Speech Intelligibility Index (CSII)

CSII was developed to estimate intelligibility in speech distorted by additive noise or bandwidth reduction. This measure is based on the SII, but uses the coherence-based (Speech to Distortion Ratio) SDR function instead of the SNR of the former. The three level CSII is presented in [14]. In the experiments performed using Hear In Noise Test (HINT) sentences both in normal hearing and hearing impaired subjects, this measure yields high correlation with the average perceptual intelligibility. The calculation of the three level CSII follows these steps:

1) *Separation:* the speech signal envelope is divided in three amplitude levels.
2) *Segmentation:* both the clean and distorted signals are split in 30 ms segments.
3) *FFT:* the short time FFT is applied to both signals.
4) *Coherence Calculation:* magnitude-squared coherence between the clean and distorted signal is calculated and summed over the entire signal.
5) *Combination:* the scores for each amplitude level are combined to compose one score for the audio signal.
6) *Mapping:* to translate the values of STOI to WRR scores, a mapping function must be used. It takes the form of

$$CSIImap = \frac{100}{1 + exp(a * CSII + b)} \tag{12}$$

in which $CSIImap$ is the predicted intelligibility score, $CSII$ is the output of the algorithm, a and b are two free parameters, adjusted according to the experiment.

IV. Experimental Results and Discussion

The methods were evaluated on a subset of the TIMIT database [18] composed of 240 speech utterances. The signals are monaural, on average 3 seconds long and are sampled at 16kHz. The experiments were conducted on speech reverberated using an implementation of the Image Source Model (ISM) method [19]. This study used 14 reverberating conditions: two different rooms, taken from [2], [9] and seven T_{60}: $0.6s, 0.8s, 1.0s, 1.2s, 1.4s, 1.6s, 1.8s$. The room sizes, source and sensor positions are summarized in Table I.

Room	Room Dimensions (m)	Source Position (m)	Receiver Position (m)
Nabelek	$[7.5 \times 6.1 \times 3.6]$	$[1, 1, 1.5]$	$[3.82, 3.82, 1.5]$
Petkov	$[20 \times 30 \times 8]$	$[10, 5, 3]$	$[10, 25, 1.8]$

TABLE I: Experimental Reverberated Conditions

The effect of reverberation and processing on speech is illustrated in Figure 1, where a speech utterance is presented, respectively: clean, i.e. non-reverberated, reverberated using the ISM algorithm in the Petkov room under reverberation time $T_{60} = 1.8$s, and then processed, respectively, with the SSS and AGC methods, explained in Section II. The processing of speech using AGC divided the audio files in 40ms lenght frames with 50% overlapping.Intelligibility measurement was performed using in a frame basis procedure with 50% overlapping in all objective measures.

The free parameters a and b of the mapping functions were calculated through nonlinear least squares fitting between the subjective intelligibility in [2] for $T_{60} = 0.8$s and 1.2s the average OIM scores in those conditions. Table II presents the parameters for each OIM, with the Subjective Intelligibility (SI) and the corresponding Predicted Intelligibility (PI) obtained after mapping.

Objective Measure	Parameters		T_{60}	SI [2]	PI
STOI	a	-4.18	0.8	86.9%	86.39%
	b	0.67	1.2	85.8%	85.35%
WSTOI	a	-4.28	0.8	86.9%	86.39%
	b	0.71	1.2	85.8%	85.53%
CSII	a	-4.55	0.8	86.9%	86.33%
	b	1.07	1.2	85.8%	85.53%

TABLE II: Mapping Parameters

A. *Intelligibility Results*

1) *STOI results*

Table III presents the average predicted intelligibility obtained using STOI, for the reverberated speech processed with SSS and AGC, as well as the unprocessed (UNP). Two room sizes were used, Nabelek $[7.5m \times 6.1m \times 3.6m]$ and Petkov $[20m \times 30m \times 8m]$. Reverberation times $T_{60} = 0.6, 0.8, 1.0, 1.2, 1.4, 1.6$ and 1.8s.

Room	T_{60}	UNP	SSS	AGC
Nabelek $[7.5 \times 6.1 \times 3.6]$	0.6	89.34%	87.59%	88.79%
	0.8	86.39%	84.74%	86.19%
	1.0	84.76%	83.20%	84.74%
	1.2	85.35%	83.77%	85.26%
	1.4	81.01%	79.66%	81.36%
	1.6	80.05%	78.77%	80.48%
	1.8	78.34%	77.18%	78.89%
Average		83.61%	82.13%	83.67%
Petkov $[20 \times 30 \times 8]$	0.6	85.53%	83.72%	85.34%
	0.8	82.45%	80.71%	82.55%
	1.0	77.89%	76.31%	78.36%
	1.2	75.40%	73.92%	76.02%
	1.4	71.11%	69.89%	72.17%
	1.6	69.71%	68.56%	70.84%
	1.8	66.03%	65.16%	67.47%
Average		75.44%	74.04%	76.11%

TABLE III: Average predicted intelligibility using STOI.

AGC results indicate the method improves speech intelligibility under intense reverberation. Room size influences the effect, as can be seen comparing the average improvement in Petkov room (0.66%) and in Nabelek room (0.07%). Note that this solution seems to be better suited to longer reverberation times. Improvement is obtained for T_{60} longer than 1.4s for Nabelek room and 0.8s for Petkov room.

SSS displays speech intelligibility degradation in all presented conditions. Contrary to AGC, the room size seems to have little effect in the performance (-1.47% intelligibility difference in Nabelek room, versus -1.40% in Petkov room). The decline on softer reverberation, though, is more accentuated than in higher reverberation (average -1.78% difference in $T_{60} = 0.6$, against -1.01% in $T_{60} = 1.8$). These results agree with the previous subjective intelligibility studies [10], [7].

2) *WSTOI Results*

The predicted intelligibility using WSTOI is summarized in Figures 2 and 3. The vertical axis represents the predicted intelligibility in %, while the horizontal axis displays the T_{60} in seconds. Unprocessed speech (UNP) is represented as the white bars, the speech processed with SSS as the grey bars and the processed by AGC as the black bars.

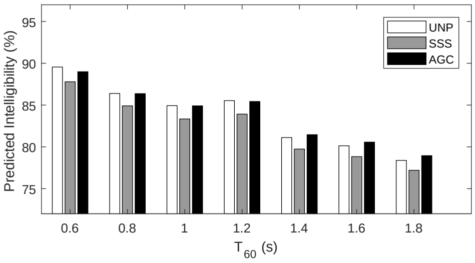

Fig. 2: Average predicted intelligibility using WSTOI in Nabelek room.

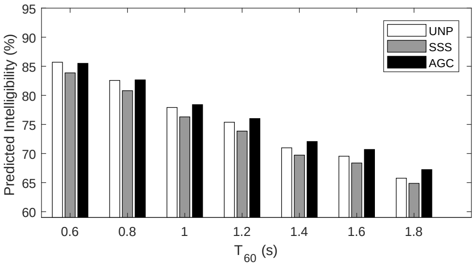

Fig. 3: Average predicted intelligibility using WSTOI in Petkov room.

As can be seen, T_{60} impacts more the speech intelligibility in a bigger room. The average difference from $T_{60} = 0.6$s to $T_{60} = 1.8$s in Nabelek room is -11.17% and in Petkov room it is -19.94%. The average intelligibility gain using AGC is 0.39%. The biggest impact is in Petkov room with $T_{60} = 1.8$s. SSS presents an intelligibility loss of an average -1.47% across all studied conditions.

3) *SRMR results*

Table IV presents the predicted intelligibility obtained with the SRMR measure.

As expected, the behavior of SSS processed speech is similar to the values obtained with other measures (an average -1.08% intelligibility difference from the unprocessed speech). The AGC processed speech, however, is underestimated (an average -0.12% loss, where all other measures accuse improvement).

4) *CSII results*

The predicted intelligibility obtained using CSII is illustrated in Figures 4 and 5. Speech is represented as with Figures 2 and 3.

This experiment indicates improvement in intelligibility of speech processed by AGC, as was also found in most other

Room	T_{60}	UNP	SSS	AGC
Nabelek $[7.5 \times 6.1 \times 3.6]$	0.6	88.56%	87.50%	87.86%
	0.8	86.14%	85.21%	85.55%
	1.0	85.08%	84.20%	84.54%
	1.2	85.35%	84.45%	84.80%
	1.4	83.15%	82.36%	82.68%
	1.6	82.74%	81.97%	82.29%
	1.8	82.06%	81.32%	81.64%
Average		84.72%	83.86%	84.19%
Petkov $[20 \times 30 \times 8]$	0.6	90.19%	88.24%	89.08%
	0.8	89.03%	87.26%	88.05%
	1.0	86.48%	86.48%	87.19%
	1.2	87.02%	85.55%	86.27%
	1.4	86.01%	84.68%	85.32%
	1.6	85.44%	84.13%	84.79%
	1.8	84.56%	83.36%	83.94%
Average		86.96%	85.67%	86.38%

TABLE IV: Average predicted intelligibility using SRMR.

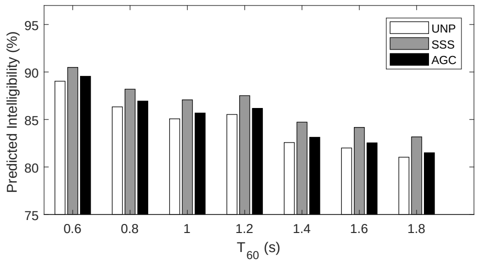

Fig. 4: Average predicted intelligibility using CSII in Nabelek room.

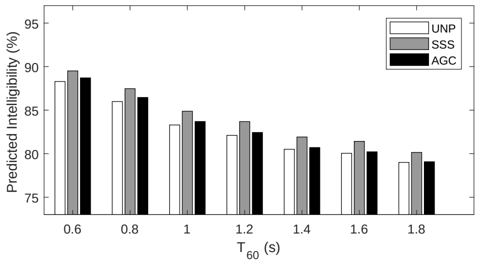

Fig. 5: Average predicted intelligibility using CSII in Petkov room.

measures (0.31% improvement). However, the pattern of SSS processed speech is the inverse (it shows improvement, where the other measures show degradation).

V. Conclusion

This work investigated the effect of two adaptive gain methods, AGC and SSS, on the intelligibility of reverberated speech under conditions never previously studied. To evaluate these methods, speech segments from the TIMIT corpus were artificially reverberated and processed, then had their intelligibility measured using four objective intelligibility measures widely known in the literature. The experiments shown AGC method improves intelligibility, mostly in bigger rooms. The effect is also more pronounced in longer T_{60}. The experiments showed the effect of room size and reverberation time in the performance of both methods. The larger the room, more pronounced is the improvement in speech intelligibility using AGC. This method also performs better in longer T_{60}. SSS processed speech was found to be less intelligible than unprocessed speech in both large and small rooms, a result that agrees with previous experiments.

References

[1] RH Bolt and AD MacDonald, "Theory of speech masking by reverberation," *The Journal of the Acoustical Society of America*, vol. 21, no. 6, pp. 577–580, 1949.

[2] Anna K Náběelek, Tomasz R Letowski, and Frances M Tucker, "Reverberant overlap-and self-masking in consonant identification," *The Journal of the Acoustical Society of America*, vol. 86, no. 4, pp. 1259–1265, 1989.

[3] Pierre J Castellano, S Sradharan, and David Cole, "Speaker recognition in reverberant enclosures," in *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*. IEEE, 1996, vol. 1, pp. 117–120.

[4] Takuya Yoshioka, Armin Sehr, Marc Delcroix, Keisuke Kinoshita, Roland Maas, Tomohiro Nakatani, and Walter Kellermann, "Making machines understand us in reverberant rooms: Robustness against reverberation for automatic speech recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 114–126, 2012.

[5] Masato Miyoshi and Yutaka Kaneda, "Inverse filtering of room acoustics," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 36, no. 2, pp. 145–152, 1988.

[6] Stephen T Neely and Jont B Allen, "Invertibility of a room impulse response," *The Journal of the Acoustical Society of America*, vol. 66, no. 1, pp. 165–169, 1979.

[7] Yusuke Miyauchi, Nao Hodoshima, Keiichi Yasu, Nahoko Hayashi, Takayuki Arai, and Mitsuko Shindo, "A preprocessing technique for improving speech intelligibility in reverberant environments: The effect of steady-state suppression on elderly people," in *Ninth European Conference on Speech Communication and Technology*, 2005.

[8] Richard C Hendriks, João B Crespo, Jesper Jensen, and Cees H Taal, "Optimal near-end speech intelligibility improvement incorporating additive noise and late reverberation under an approximation of the short-time sii," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 5, pp. 851–862, 2015.

[9] Petko N Petkov and Yannis Stylianou, "Adaptive gain control for enhanced speech intelligibility under reverberation," *IEEE signal processing letters*, vol. 23, no. 10, pp. 1434–1438, 2016.

[10] Takayuki Arai, Nao Hodoshima, and Keiichi Yasu, "Using steady-state suppression to improve speech intelligibility in reverberant environments for elderly listeners," *IEEE transactions on audio, speech, and language processing*, vol. 18, no. 7, pp. 1775–1780, 2010.

[11] Cees H Taal, Richard C Hendriks, Richard Heusdens, and Jesper Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, 2011.

[12] Leo Lightburn and Mike Brookes, "A weighted stoi intelligibility metric based on mutual information," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 5365–5369.

[13] Tiago H Falk, Chenxi Zheng, and Wai-Yip Chan, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1766–1774, 2010.

[14] James M Kates and Kathryn H Arehart, "Coherence and the speech intelligibility index," *The journal of the acoustical society of America*, vol. 117, no. 4, pp. 2224–2237, 2005.

[15] Sadaoki Furui, "On the role of spectral transition for speech perception," *The Journal of the Acoustical Society of America*, vol. 80, no. 4, pp. 1016–1025, 1986.

[16] Kostas Kokkinakis and Philipos C Loizou, "The impact of reverberant self-masking and overlap-masking effects on speech intelligibility by cochlear implant listeners (l)," *The Journal of the Acoustical Society of America*, vol. 130, no. 3, pp. 1099–1102, 2011.

[17] Hanna Järveläinen and Matti Karjalainen, "Reverberation modeling using velvet noise," in *Audio Engineering Society Conference: 30th International Conference: Intelligent Audio Environments*. Audio Engineering Society, 2007.

[18] John S Garofolo, Lori F Lamel, William M Fisher, Jonathon G Fiscus, and David S Pallett, "Darpa timit acoustic-phonetic continous speech corpus cd-rom. nist speech disc 1-1.1," *NASA STI/Recon technical report n*, vol. 93, 1993.

[19] Eric A Lehmann and Anders M Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 269–277, 2008.