INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Generalizing Steady State Suppression for Enhanced Intelligibility under Reverberation

Petko N. Petkov and Yannis Stylianou

Toshiba Research Europe Ltd., Cambridge Research Laboratory, Cambridge, United Kingdom
{petko.petkov,yannis.stylianou}@crl.toshiba.co.uk

Abstract

Speech intelligibility in reverberant environments decreases due to overlap-masking. Unlike additive noise, the masking signal is not independent from the information bearing signal. A mathematical framework for intelligibility-enhancing signal modification prior to presentation in reverberant environments is presented in this paper. The optimal solution generalizes steady state suppression and adjusts the short-term signal power as a function of late reverberation power and signal importance. The signal modification operates in a full-band setting and preserves the time scale of the unmodified signal. Gain smoothing based on an adaptive rate-of-change constraint reduces processing artifacts and enhances performance. Subjective validation shows that the proposed method effectively reduces the impact of overlap-masking. Speech intelligibility at a reverberation time of 1.8 s was improved significantly compared to unmodified and steady-state-suppressed speech.

Index Terms: speech intelligibility, reverberation, speech modification, power dynamics recovery

1. Introduction

The overlap-masking effect, i.e., the simultaneous observation of a large number of delayed and attenuated copies of an acoustic signal, degrades the intelligibility of speech in reverberant environments [1]. Reverberation consists of early reflections (ER) arriving within a small window after the direct (or shortest propagation path) sound and late reverberation (LR) [2]. While ER are distinct and depend on the hall geometry and the positions of the speaker and the listener, LR is diffuse due to the multitude of times each signal copy has been reflected. LR has been identified as the primary cause of intelligibility degradation under reverberation [1, 3].

A number of intelligibility enhancing speech modifications for reverberation have been proposed. Inverse filtering aims to cancel the reverberation effect but lacks spatial robustness and suffers from limitations related to the invertibility of the room impulse response (RIR) [4, 5]. Modulation enhancement to offset the loss of modulation depth caused by reverberation is proposed in [6]. Based on performance figures, practical application of this methodology requires further improvements.

Reducing the impact of overlap-masking on transient portions of the signal by steady state suppression (SSS) is discussed, e.g., in [7, 8, 9]. The approach is motivated by the importance of sound transitions. Evidence of the method effectiveness is primarily based on syllable recognition tasks [7, 10]. Intelligibility degradation is reported in [9, 11]. The inconsistent performance is likely caused by insufficient adaptation to the speech statistics and the environment, as well as distortion introduced by rapid gain fluctuations.

Local and global time-scale modifications are significantly more effective in recovering speech transients. Zero-padding in the steady state is proposed in [12] while fixed time-scaling is evaluated in [13]. These methods do not adapt to the specific conditions and apply a user-defined modification depth. Use of fixed-duration pauses without linguistic motivation is considered in [14, 9]. An approach to using linguistically-motivated pauses with context-adaptive duration is studied in [15]. The intelligibility gain achieved by these methods comes at the cost of a reduced information transfer rate. The combination of time-scale modification and signal power adjustment is likely to attract more attention in the future.

Recent years have seen an increase in interest for optimization-based methods. Perceptual distortion is minimized for the parameters of a spectral gain modification in [16, 17]. A speech intelligibility index (SII) [18] based measure is optimized in [19]. Local SII optimization by spectral shaping and dynamic range compression is studied in [20]. The methods listed above address reverberation exclusively in combination with noise while focusing on mild-to-moderate reverberation conditions.

In this paper we revisit the idea of suppressing signal power in the steady state [7] to reduce masking of sound transitions. A mathematical framework is used, where by optimization a short-term full-band power correction is obtained. The solution depends on the LR power and the signal importance. An extended theoretical analysis of the methodology is presented in [21]. Once the locally-optimal gain is computed, it is smoothed to reduce rapid gain fluctuations. The proposed approach generalizes steady state suppression (SSS) [7] and effectively improves intelligibility for meaningful sentences. Performance is validated with a listening test.

The remainder of this paper is organized as follows. Theory is presented in Section 2. System design is discussed in Section 3. Experimental results are given in Section 4 followed by conclusions.

2. Theoretical foundations

The modification of a speech-in-noise distortion criterion for the specifics of the reverberation problem is presented in Section 2.1. Calibration of the optimal input-output (IO) signal power mapping with respect to LR power is summarized in Section 2.2. Dependence of the output power on local signal importance is discussed in Section 2.3.

2.1. Power gain optimization

A distortion criterion quantifying the effect of additive noise on power dynamics was proposed in [22] and used for intelligibil-

http://dx.doi.org/10.21437/Interspeech.2016-1026

ity enhancing speech modification. Raising the LR (noise in [22]) power, in the gain penalty of the criterion, to an exponent larger than one introduces a new functionality. The distortion criterion, using power two for the exponent, is given by:

$$\eta = \int_{\alpha}^{\beta} \left(\frac{1}{x} \left(y + l - x \frac{dy}{dx} \right)^2 + \lambda l^2 \frac{y}{x} \right) f_X\left(x|b\right) dx, \quad (1)$$

where x, y and l are the instantaneous powers of unmodified speech, modified speech and LR respectively, λ is a Lagrange multiplier, $f_X\left(x|b\right)$ is the probability density function of the Pareto distribution [23] with shape parameter b, and α and β define the optimal operating range. The first additive term is a distortion measure and the second is the power gain penalty. With an increase in l, the penalty gradually outweighs the distortion, and inverses the modification direction.

Optimizing (1) based on calculus of variations [24], a closed-form power mapping $y\left(x\right)$ is obtained as:

$$y\left(x\right) = c_1 x + c_2 x^b + \frac{l}{2b}\left(l\lambda - 2b\right), \quad (2)$$

where c_1 and c_2 are constants. The solution is readily verified to be a minimizer [22, 25]. The boundary conditions used to determine c_1 and c_2 are:

$$y\left(\alpha\right) = \alpha, \quad (3) \qquad y'\left(\psi\right) = \rho, \quad (4)$$

where $y' = \frac{dy}{dx}$, $\rho = \varsigma^l$, $\varsigma \in (0,\ 1)$ is a small positive number and $\psi \to \infty$ is a large positive number. Dependence of the derivative on the LR power l is introduced to ensure that in the absence of reverberation ($y'\left(\psi\right) = 1$), the signal remains unchanged. In the presence of reverberation, $\rho \to 0$ is effectively a constant. c_1 and c_2 are obtained by solving the linear system formed by equations (3) and (4).

2.2. IO power map calibration

Practical use of (2) requires that the critical LR power $\tilde{l}$, which causes an inversion in modification direction, is known in advance. Facilitated by the choice for the LR power exponent in (1), we calibrate the mapping such that for $\lambda = \tilde{\lambda}, l = \tilde{l}$ induces maximum boosting $y = \beta$. Starting from:

$$y\left(x = \beta|\lambda, l\right) = \beta \quad (5)$$

and grouping all terms along the powers of l, produces a quadratic form. Solving the single-root condition of this quadratic form for λ identifies the multiplier $\tilde{\lambda}$ as:

$$\tilde{\lambda} = \frac{b}{2\left(1-\rho\right)} \frac{\beta^b - \alpha^b - \left(\beta - \alpha\right) b\psi^{b-1}}{\alpha^b\beta - \alpha\beta^b}. \quad (6)$$

The LR power $\tilde{l}$ at which the inversion occurs is the single root of the quadratic form:

$$\tilde{l} = {}^{b}\!/_{\tilde{\lambda}}. \quad (7)$$

A family of IO power mappings is shown in Figure 1. The crossing point between the optimal mapping and $y = x$ for $x > \alpha$ defines the maximum boosting power (MBP). The mapping is guaranteed to be monotonically increasing on $x \in (\alpha, \psi)$ when $y'\left(\psi\right) \to 0$ and MBP exceeds α.

The IO mapping is not lower-bounded in general. The monotonic increase property can be violated for $\lambda = \tilde{\lambda}$ given a sufficiently large l. A particular MBP $\nu \in (\alpha,\ \beta]$ can be

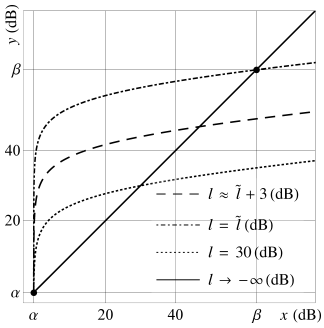

Figure 1: Input-output power mappings for a set of LR power values and $\lambda = \tilde{\lambda}$.

achieved independent of the value of l, $l > 0$. The multiplier λ_ν that ensures this behavior is derived by substituting β with ν in (5) and solving for λ:

$$\lambda_\nu = \frac{2b}{l^2} \frac{\left(\rho - 1\right)\left(\alpha^b\nu - \alpha\nu^b\right)}{\nu^b - \alpha^b - b\left(\nu - \alpha\right)\psi^{b-1}} + \frac{2b}{l}. \quad (8)$$

2.3. Signal importance and optimal gain

We refer to signal importance as an indicator of how non-stationary the signal is. Full context awareness requires dependence on the degree of signal importance in the IO power mapping. This dependence can be introduced through the multiplier λ. The objective is to permit more boosting (in case of $l \leq \tilde{l}$) and less suppression ($l > \tilde{l}$) when the signal is less stationary. Inversely, less boosting ($l \leq \tilde{l}$) and stronger suppression ($l > \tilde{l}$) is targeted when the signal is more stationary.

One approach to achieve the desired functionality is to establish a monotonic dependence of MBP on the signal importance, which we denote by ξ. The sigmoid:

$$q\left(\Theta \mid s,\ H,\ L\right) = \frac{1 - e^{-s\Theta}}{1 + e^{-s\Theta}}\left(H - L\right) + L,\ \Theta > 0, \quad (9)$$

with slope s, is used here to map ξ to an MBP ν_ξ bounded by $L = \alpha$ and $H = \beta$ in log domain. The multiplier λ_{ν_ξ} that achieves MBP ν_ξ is identified from (8).

For $l \leq \tilde{l}$, the effective multiplier:

$$\lambda = \max\left(\lambda_{\nu_\xi},\ \tilde{\lambda}\right) \quad (10)$$

prevents boosting beyond MBP ν_ξ. For $l > \tilde{l}$:

$$\lambda = \lambda_{\bar{\nu}} \quad (11)$$

$\left(\log\left(\bar{\nu}\right) = q\left({}^{\lambda_{\nu_\xi}}\!/_{\tilde{\lambda}} \mid s,\ \log\left(\nu_\xi\right),\ \log\left(\nu_\alpha\right)\right)\right)$ ensures MBP $\bar{\nu} \in (\alpha,\ \nu_\xi)$.

3. System design

An operation diagram of the proposed adaptive gain control (AGC) system is illustrated in Figure 2. Overlapping frames are extracted from the input speech signal and labeled according to their importance. A late reverberation model predicts LR power. The optimal output power is computed given the input power, the LR power and the signal importance. Frame-based estimates are used to approximate instantaneous power and importance. The output power is smoothed to reduce artifacts. The short-term signal is scaled, reflecting the target power, and added to the buffer. Importance estimation, LR modeling and gain smoothing are summarized below.

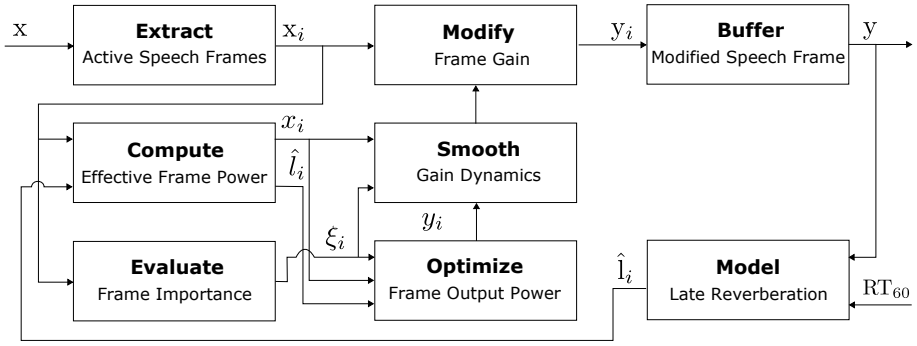

Figure 2: Operation diagram of the proposed system.

3.1. Frame importance estimator

We propose a causal frame importance estimator based on measuring the normalized distance of the Mel frequency cepstral coefficients (MFCCs) in adjacent frames:

$$\xi_i \;=\; \frac{\|\mathrm{m}_i - \mathrm{m}_{i-1}\|}{\|\mathrm{m}_i\| + \|\mathrm{m}_{i-1}\|},\; \xi_i \;\in\; [0,\, 1]\,, \tag{12}$$

where m_i is the MFCC set from frame i. High degree of similarity in the feature domain indicates that the signal is stationary and translates to low importance. The estimator behavior is shown in Figure 3.

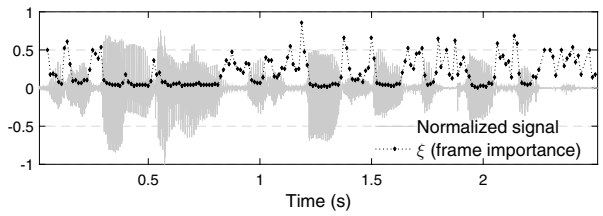

Figure 3: Active-frame importance estimates. Estimator fluctuation towards the end of the waveform is caused by on/off switching in the voice activity detector.

3.2. Late reverberation model

Let τ (in seconds) denote the boundary between early reflections and LR [2], as measured from the arrival of the direct-path sound. A simple model assuming the exponential decay of LR power with time and a constant RT_{60} over frequency is used here [26]. The LR part of the RIR is modeled as a pulse train $\iota\,[k]$, which is amplitude modulated by an exponentially decaying function:

$$\tilde{h}\,[k] = \iota\,[k]\,10^{-3\frac{k}{\mathrm{RT}_{60} f_s}}, \tag{13}$$

where f_s is the sampling rate. The energy of the modulated pulse train is equalized to the energy of the LR part of RIR calculated from a measurement. The approximate LR waveform $\hat{\mathrm{l}}$ is given by the convolution:

$$\hat{\mathrm{l}}\,[k] = \sum_{n=1}^{(\mathrm{RT}_{60}-\tau)f_s} \tilde{h}\,[\tau f_s + n]\,\mathrm{y}\,[k - \tau f_s - n]\,. \tag{14}$$

A sample-based LR power estimate $\hat{l}$ is computed from $\hat{\mathrm{l}}$. Given that full-band operation is considered at present, the over-adjustment of signal power is reduced by working with the LR power from the spectral range dominated by the speech signal. A set of non-overlapping frequency bands, linearly-distributed on a Mel scale, is used for the purpose. The effective LR power is computed from the bands with the largest contributions that contain at least 90 % of the speech power.

3.3. Gain smoothing

Rapid gain fluctuation creates discontinuities with negative impact on intelligibility. Limiting the rate of change smooths the locally-optimal gain without smearing frame importance. The smoothed gain $\check{g}$ is constrained by:

$$D < \check{g}_i \leq U^{g_i^{\phi_i}}, \quad g_i^2 = {}^{y_i}\!/_{x_i} \tag{15}$$

where D and U are constants, and ϕ_i may vary depending on the speech properties. The particular upper limit permits stronger boosting for low-power signal portions. The effective rates converge to their limits with ξ:

$$\{u_i,\; d_i\} \equiv \left\{ q\left(\xi_i \,|s, U^{g_i^{\phi_i}}, 1\right),\; q\left(\xi_i \,|s, 1, D\right) \right\}, \tag{16}$$

where the smooth mapping from (9) was used for convenience. The smooth signal-domain gain is:

$$\check{g}_i = \begin{cases} \min\left(u_i,\; g_i\right) & \text{if } g_i > 1 \\ \max\left(d_i,\; g_i\right) & \text{if } g_i \leq 1. \end{cases} \tag{17}$$

4. Evaluation

Reverberation is simulated using a source-image method generated RIR [27]. The assumed hall dimensions are $20 \times 30 \times 8$ m, with speaker and listener locations $\{10,\; 5,\; 3\}$ and $\{10,\; 25,\; 1.8\}$ m respectively. For convenience, propagation delay and attenuation are normalized to the direct sound. The training data used to fit $f_X\,(x)$), and determine α and β, is a British English recording of [28] comprising 720 sentences.

Table 1: System parameter values.

Frame duration: 25 ms	Frame overlap: 50 %
m: MFCCs orders 1-to-12	$\tau = 0.05$, [2]
α: min frame power (all data)	$\varsigma = 0.001$
β: max frame power (all data)	$\psi = \beta^4$
Pulse density in ι: 4000 s^{-1}	$s = 20$
$D = 0.15$	$U = 1.15$
$\phi_i \in \{1/3, 1/6\}$	

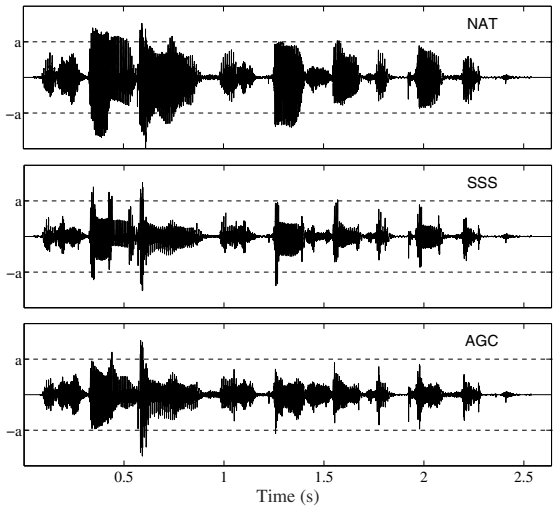

Figure 4: Signal waveforms for the test sentence from Figure 3 preprocessed for presentation at $\mathrm{RT}_{60} = 1.8$ s.

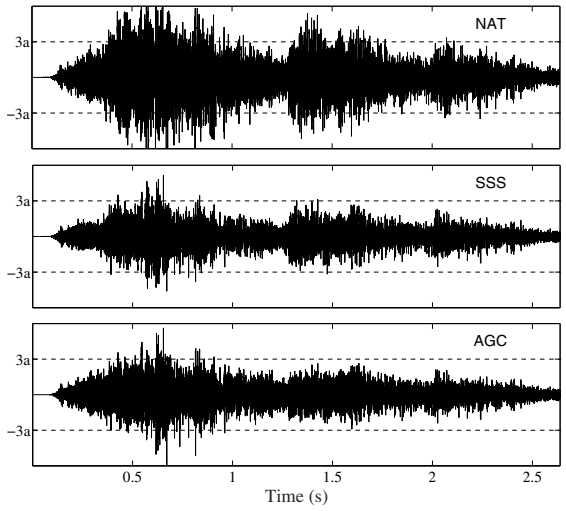

Figure 5: Processed reverberant waveforms, for the test signals from Figure 4, at $\mathrm{RT}_{60} = 1.8$ s.

Values of the operational parameters from the proposed system design are listed in Table 1. In addition, effective LR power was determined based on a ten-band set-up. The smaller value for ϕ was used when the largest speech power contribution came from the range above 2.6 kHz (band six for computing the effective LR power). This reduced a high-frequency artifact likely caused by the flat response of the simple reverb simulator.

Figure 4 shows the output waveforms from the three methods included in the evaluation for the sentence ("*The swan dive was far short of perfect.*") processed for presentation at $\mathrm{RT}_{60} = 1.8$ s. The unmodified signal is denote by NAT. The reference system is an in-house implementation of [7]. The envelope profile for SSS is consistent with, e.g., [13, 29]. The corresponding reverberant waveforms are shown in Figure 5. In both figures $a > 0$ is the same constant.

We observe, from Figure 3 and Figure 4, that low importance regions undergo gain suppression for both SSS and AGC. The gain reduction rate is clearly faster for SSS, which causes audible artifacts. On average, as measured over the 170 sentences from set 39 through to set 55 in [28], both AGC and SSS reduce signal power by approximately 60 %. This number is also an indication of a significant reduction in reverberation power. In the absence of reverberation, the AGC-modifed waveform converges to the NAT waveform.

AGC has a low algorithmic delay due to the causality of the importance estimator. In contrast, SSS [7] uses a look-ahead window [30]. The method complexity is low, with LR waveform computation (eq. (14)) as the most demanding task. Real-time processing is achieved in Matlab [31] by accounting for the sparsity of $\tilde{h}$ from eq. (13). Significant decrease in complexity, considering the particular LR power estimator, can be achieved by pulse density reduction and pulse train truncation.

A listening test with five native English speakers (average age 29) was conducted to compare the intelligibility of unmodifed, SSS-modified and AGC-modified speech. The participants, recruited from the affiliated research facility, did not report any hearing impairments. Modified utterance power was equalized to facilitate comparison. The material was presented diotically, in a silent room, using a pair of Audio-technica ATH-M50x headphones.

A preliminary session using sets 39 and 40 from [28] familiarized the subjects with the task and the test interface. The test material comprised sets 41 to 55 (150 sentences in total). Each method was assigned a macro set of five sets. The allocation of macro-set to system and the system presentation order were randomly selected for each listener. After hearing a sentence once, the listener was prompted to type its content. A word recognition rate (WRR) was computed for each sentence as the ratio of correctly-identified to the total number of keywords [32]. The average (over macro sets and subjects) recognition rates and standard errors for each method are shown in Figure 6. Five subjects were sufficient to measure significant intelligibility gain for AGC over NAT ($p < 0.05$, Student's t test). SSS degrades intelligibility insignificantly compared to NAT, a result obtained independently, e.g., in [9].

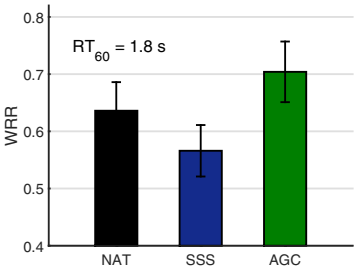

Figure 6: Average word recognition rates (WRR).

5. Conclusions

A mathematical framework optimizing the full-band signal gain as a function of input power, late reverberation power and signal importance produces a context-aware generalization of steady-state suppression. Continuous and gradual adaptation combined with adaptive gain smoothing result in artifact-free processing. The proposed method achieves significant intelligibility gain over natural speech and steady-state suppression.

6. References

[1] R. H. Bolt and A. D. MacDonald, "Theory of Speech Masking by Reverberation," *J. Acoust. Soc. Am*, vol. 21, no. 6, pp. 577–580, 1949.

[2] T. Yoshioka, A. Sehr, M. Delcroix, K. Kinoshita, R. Maas, T. Nakatani, and W. Kellerman, "Making Machines Understand Us in Reverberant Rooms," *IEEE Sig. Proc. Mag.*, vol. 29, no. 6, pp. 114–126, 2012.

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

[4] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Am.*, vol. 66, no. 1, pp. 165–169, 1979.

[5] A. Mertins, T. Mei, and M. Kallinger, "Room Impulse Response Shortening/Reshaping with Infinity- and p-Norm Optimization," *IEEE Trans. Audio, Speech and Lang. Proc.*, vol. 18, no. 2, pp. 249–259, 2010.

[6] A. Kusumoto, T. Arai, K. Kinoshita, N. Hodoshima, and N. Vaughan, "Modulation Enhancement of Speech as a Preprocessing for Reverberant Chambers with the Hearing-Impaired," *Speech Communication*, vol. 45, pp. 101–113, 2005.

[7] 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. Am.*, vol. 119, pp. 4055–4064, 2006.

[8] M. Tsuji, T. Arai, and K. Yasu, "Preprocessing using consonat emphasis and vowel suppression for improving speech intelligibility in reveberant environments," *J. Acoust. Soc. Japan*, vol. 69, no. 4, pp. 179–183, 2013.

[9] F. Fuhrmann, K. Dobbler, F. Pokorny, and F. Graf, "A modular system for improving speech intelligibility under extreme acoustic conditions: Subjective evaluation of parameter influence," in *Proc. Forum Acusticum*, 2014.

[10] 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 and Lang. Proc.*, vol. 18, no. 7, pp. 1775–1780, 2010.

[11] N. Hayashi, T. Arai, N. Hodoshima, Y. Miyauchi, and K. Kurisu, "Steady-state pre-processing for improving speech intelligbility in reverberant environemtns: Evaluation in a hall with an electrical reverberator," in *Proc. Interspeech*, 2005, pp. 1741–1744.

[12] T. Arai, "Padding zeros into steady-state portions of speech as a preprocess for improving intelligibility in reverberant environments," *Acoust. Sc. & Tech.*, vol. 26, no. 5, pp. 459–461, 2005.

[13] Y. Nakata, Y. Murakami, N. Hodoshima, N. Hayashi, Y. Miyauchi, T. Arai, and K. Kurisu, "The Effects of Speech-Rate Slowing for Improving Speech Intelligibility in Reverberant Environments," The Institute of Electr., Inf. & Comm. Eng., Tech. Rep., 2006.

[14] F. Sattar, M. Nilsson, and I. Claesson, "Segmentation and its real-world applications in speech processing," in *Intern. Symposium on Signal Processing and its Applications*, 2008.

[15] P. N. Petkov, N. Braunschweiler, and Y. Stylianou, "Automated pause insertion for improved intelligibility under reverberation," in *Proc. Interspeech*, 2016.

[16] J. B. Crespo and R. C. Hendriks, "Speech Reinforcement in Noisy Reverberant Environments Using a Perceptual Distortion Measure," in *Proc. ICASSP*, 2014, pp. 910–914.

[17] ——, "Speech Reinforcement with a Globally Optimized Perceptual Distortion Measure for Noisy Reverberant Channels," in *Proc. IWAENC*, 2014, pp. 89–93.

[18] American National Standard, "Methods for the Calculation of the Speech Intelligibility Index," 1997.

[19] 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 Trans. Audio, Speech and Lang. Proc.*, 2015.

[20] H. Schepker, D. Hülsmeier, J. Rennies, and S. Doclo, "Model-based integration of reverberation for noise-adaptive near-end listening enhancement," in *Proc. Interspeech*, 2015, pp. 75–79.

[21] P. N. Petkov and Y. Stylianou, "Adaptive Gain Control for Enhanced Speech Intelligibility under Reverberation," *Accepted for publ. in IEEE Sig. Proc. Letters*, 2016.

[22] P. N. Petkov and W. B. Kleijn, "Spectral Dynamics Recovery for Enhanced Speech Intelligibility in Noise," *IEEE Trans. Audio, Speech and Lang. Proc.*, vol. 23, no. 2, pp. 327–338, 2015.

[23] B. C. Arnold, *Pareto Distributions*. International Co-operative Publishing House, 1983.

[24] R. Courant and D. Hilbert, *Methods of Mathematical Physics*. John Wiley & Sons, 1953, vol. 1.

[25] B. Chachuat, "Nonlinear and Dynamic Optmization: From Theory to Practice," Automatic Control Laboratory, EPFL, Switzerland, 2007.

[26] M. Karjalainen and H. Järveläinen, "Reverberation Modeling Using Velvet Noise," in *Proc. AES 30th Int. Conf.*, 2007.

[27] J. B. Allen and D. A. Berkley, "Image method for effciently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.

[28] "IEEE Recommended Practice for Speech Quality Measurements," *IEEE Trans. Audio and Electroacoustics*, vol. 17, pp. 225–246, 1969.

[29] T. Arai, Y. Nakata, N. Hodoshima, and K. Kurisu, "Decreasing speaking rate with steady-state suppression to improve speech intelligibility in reverberant environments," *Acoust. Sc. & Tech.*, vol. 28, no. 4, pp. 282–285, 2007.

[30] S. Furui, "On the role of spectral transition for speech perception," *J. Acoust. Soc. Am*, vol. 80, no. 4, pp. 1016–1025, 1986.

[31] http://www.mathworks.com, "Matlab2014b," 2014.

[32] M. Cooke, C. Mayo, C. V. Botinhao, Y. Stylianou, B. Sauert, and Y. Tang, "Evaluating the Intelligibility Benefit of Speech Modifications in Known Noise Conditions," *Speech Communication*, vol. 55, pp. 572–585, 2013.