# A NEW BLIND DEREVERBERATION ALGORITHM USING CHANNEL SELECTION

*Amaro A. de Lima**, *Thiago de M. Prego**, *Sergio L. Netto*

Program of Electrical Engineering, COPPE, Federal University of Rio de Janeiro, RJ, Brazil.
e-mail: $\{$*amaro.lima, thiago.prego, sergioln*$\}$@*smt.ufrj.br*

*Federal Center for Technological Education Celso Suckow da Fonseca (CEFET-RJ),
Nova Iguaçu, RJ, Brazil.

## ABSTRACT

This paper addresses the problem of reducing the reverberation effect from speech signals, which is known as dereverberation. The main idea is to modify a dereverberation algorithm based on ideal channel selection (ICS) from an algorithm with reference to a blind algorithm. The channel selection technique performs a comparison between clean and degraded speech signals to decide the channel selectivity and a blind spectral subtraction technique (SS) was used in order to estimate the clean speech for ICS. Thus, the combination of SS and ICS generated a blind dereverberation technique named proposed approach version 1, which overcame the performance of the original ICS techniques in about 35% of the reverberation quality assessment ($Q_m$). Furthermore, a different concept was adopted making the channel selection threshold varies with input speech signal instead of having a fixed value, leading to a slightly different dereverberation algorithm called proposed approach version 2, which reached $Q_m$ performance improvement of 38% compared to ICS.

***Index Terms***— Dereverberation, Channel selection, Perceptual quality assessment.

## 1. INTRODUCTION

The reverberation effect can depreciate speech intelligibility and quality, affecting mainly the performances of speech/speaker recognition and hearing aids systems, and thus requiring the use of speech enhancement techniques.

This paper is based on ideal channel selection (ICS) [1] algorithm, which was designed to enhance speech signals degraded by reverberation and additive noise. The alleged advantage of this technique is that it does not need to perform the room impulse response (RIR) inversion in order to obtain a dereverberated speech signal, since RIRs of highly reverberant rooms have thousands of filter taps, making their inversion computationally expensive and reduce the consonant errors in intelligibity. The concept of the algorithm consists in comparing the frequency bins (channels) ratio of the windowed reverberant and clean speech signals, i.e., the signal-to-reverberant ratio (SRR), with a threshold. If the channel SRR is greater than the threshold, the degraded speech signal channel energy sample is selected for the output signal, otherwise it is set to zero. The aim of the threshold is to retain the components originated from early reflections and discard the components of late reverberation, due to the fact that the first are known to improve and the latter causes deterioration in speech intelligibility, since it smears the temporal envelope the speech signal.

In order to emulate the clean speech to be applied in ICS algorithm, the spectral subtraction (SS) block proposed in [2] was used, generating the spectral subtracted speech signal, which can be considered as an estimate of the clean speech signal. The SS algorithm differs from the tradition SS techniques, because the first aims to reduce the reverberation effect and the traditional ones are applied for noise reduction.

This paper is organized as follows: In Section 2, the original SS [2] and ICS [1] algorithms are detailed, with focus given on the steps to be changed in the proposed procedure; Section 3 describes two blind dereverberation algorithms propositions. One is based on the direct combination of SS and ICS algorithms using a fixed threshold, while the other uses a mapping function to relate the blind reverberation time to the threshold to be used for a given speech signal according to the reverberation quality assessment for the given speech signal.; Section 4 describes the training and test databases employed in this work, each one comprising 100 speech signals with distinct reverberation levels; Section 5 is divided into two parts, where the first presents the procedures adopted to generate the mapping function for the second proposed algorithm, and the second shows the dereverberation techniques performances for training and test databases, observing reverberation quality associated measures, as quality assessment of reverberated speech $Q_m$ [4], reverberation time ($T_{60}$) [6], room spectral variance ($\sigma_r^2$) [7] and direct-to-reverberant energy ratio ($R$) [8, 9]; Finally, conclusions concerning the relative performances increase for the dereverberation algorithms and the efficiency of the proposed approaches are addressed in Section 6.

## 2. ALGORITHMS

### 2.1. Spectral subtraction

The spectral subtraction (SS) algorithm is proposed in [2], which describes a technique for reducing the effects of late reverberation based on an adaptive approach presented in [3].

The SS algorithm considered in this work is exclusively reverberation reduction, differently from the the tradition SS algorithms, which are intended to reduce the background noise.

Let $S_y(k,m) = |S_y(k,m)|e^{j\varphi_y(k,m)}$ be the $m$-th frame of the Short-time Fourier transform (STFT) of the degraded speech signal $y(n)$. Also let $\rho$ be the length in frames of the early reflection, commonly considered to be around 50 ms and $\gamma$ be the scaling factor that establishes the relative strength of the late impulse components after the inverse filtering and $w(m)$ be an asymmetrical smoothing window based on the Rayleigh distribution.

Following these definitions, the late-reverberation power spectrum can be modeled by the convolution

$$|S_l(k,m)|^2 = \gamma w(m-\rho) * |S_y(k,m)|^2, \tag{1}$$

and the power spectrum of the early impulse components is given by

$$|S_s(k,m)|^2 = \max\left[1 - \frac{|S_l(k,m)|^2}{|S_y(k,m)|^2}, \epsilon\right], \tag{2}$$

where the auxiliary parameter $\epsilon$ keeps $|S_s(k,m)|^2$ from becoming negative or too close to zero. Finally, the magnitude spectrum of the SS estimate $\hat{x}(n)$ of the clean speech signal $x(n)$ can be determined as

$$|S_{\hat{x}}(k,m)| = \sqrt{|S_y(k,m)|^2 \times |S_s(k,m)|^2}, \tag{3}$$

and the spectrum of $\hat{x}(n)$ is estimated as

$$S_{\hat{x}}(k,m) = |S_{\hat{x}}(k,m)|e^{j\varphi_y(k,m)}. \tag{4}$$

### 2.2. Ideal channel selection

The ideal channel selection algorithm was proposed in [1] and consists in applying the clean and degraded speech signals into a kind of time-frequency mask to reduce the effects of reverberation and noise. Initially, the STFT representations $S_x(k,m)$ and $S_y(k,m)$ of $x(n)$ and $y(n)$ are obtained using a Hamming window of 20 ms with 10 ms overlap with $K$ channels. Channels are all bins corresponding to the digital frequencies in the interval $[0,\pi)$ rad/sample. Then a procedure to select or discard the spectral magnitude of $S_y(k,m)$ based on speech-to-reverberant ratio (SRR) is adopted, where the SRR is given by $SRR_{k,m} = 10\log_{10}\left(\frac{S_x(k,m)}{S_y(k,m)}\right)$. If $SRR_{k,m}$ is greater than a threshold $\tau$, then the spectral magnitude $|S_y(k,m)|$ is selected, otherwise it is discarded. Originally, $\tau = -8\ dB$ is appropriated to noise plus reverberation scenario. Finally, the frequency domain representation is converted to time domain by using Inverse STFT (ISTFT), resulting in the ICS estimate $\tilde{x}(n)$ of the clean speech signal.

## 3. PROPOSED ALGORITHMS

The first proposed algorithm entitled proposed approach version 1 combines both SS and ICS algorithms as depicted in Figure 1. The degraded speech signal $y(n)$ is used as input for the SS algorithm, which generates as output an estimate $\hat{x}(n)$ of the clean speech. This estimate of the clean speech is then used together with the degraded speech signal as inputs for the ICS algorithm, resulting in the SS-ICS estimate $\tilde{x}(n)$ of the clean speech signal.

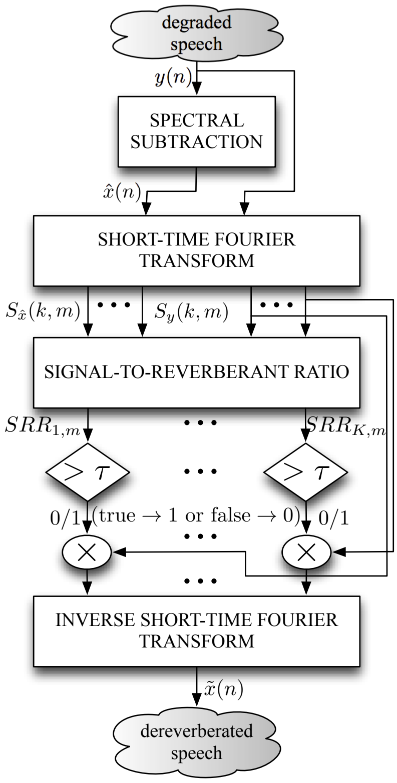

**Fig. 1**. Proposed algorithm - version 1.

The second proposed algorithm entitled proposed approach version 2, shown in Figure 2, extends the proposed approach version 1 by applying an estimation technique to adapt the threshold value of the ICS algorithm, depending on the reverberation time of the degraded speech. The blind reverberation time $T_{60}^b$ is estimated using only the input speech $y(n)$ through the use of the algorithm described in [5].

The estimated $T_{60}^b$ is then applied to a piecewise cubic Hermite interpolating polynomial (PCHIP) mapping, which maps each two points of the training data by a cubic hermite polynomial, associating the reverberation time to the most appropriated threshold $T_{hr}$, in order to maximize the quality of the dereverberated speech $\tilde{x}(n)$.

## 4. DATABASE

In this work we employed the so-called New Brazilian Portuguese (NBP) database [2, 4, 5], which composed of three reverberation scenarios:

- Artificial reverberation: This approach employed 6 artificially generated RIRs, where the early reflections were modeled via the image method, and the late reverberation used the feedback delay network

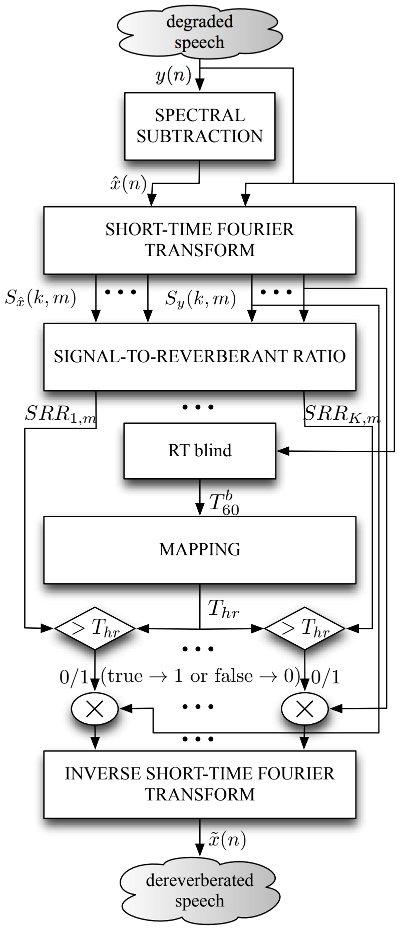

**Fig. 2**. Proposed algorithm - version 2.

method and a modified version of Gardner's method, for emulating the lower and higher reverberation times, respectively. The estimated RTs for each RIR were $\{200, 290, 390, 470, 570, 660\}$ ms.

- Natural reverberation: This approach used the RIRs determined from the direct recordings in 4 different types of rooms, with several source-microphone distances for each room, as detailed in [10], making a total of 17 RIRs. The average measured reverberation time for the 4 rooms were $\{120, 230, 430, 780\}$ ms.
- Real reverberation: In this case, the degraded signals were directly played/recorded in the 7 distinct rooms, employing at least 3 different source-microphone distances, yielding a total of 27 RIRs with average RTs in the range of $\{140, 390, 570, 650, 700, 890, 920\}$ ms.

To generate the complete NBP database, we employed 4 anechoic speech signals (2 from a male speaker and 2 from a female speaker), containing two short Brazilian-Portuguese sentences separated by approximately 1.7 s, giving an 8.4-s average duration for the entire database. These anechoic signals were used to generate 24 speech signals with artificial reverberation, 68 with real reverberation, and 108 with natural reverberation effects, making a total of 200 speech signals, all of them sampled at $F_s = 48$ kHz. The perceived quality of all speech signals was assessed through an absolute category rate (ACR) MOS test performed by 30 listeners. The whole database was sorted by the subjective MOS order. The training and test databases were separated using odd and even indexes of the sorted speech signals, respectively, resulting in 2 sub-databases of 100 reverberant speech signals each. The training database $A_1$ was used for parameter optimization, and the test database $A_2$ was used to validate the resulting system's performance, only for the proposed algorithm version 2 described in Section 3.

## 5. EXPERIMENTAL RESULTS

The experimental results is divided into two parts, where the first considers the mapping block of Section 3 and the second performs a comparison of all related techniques against the proposed approaches analyzing appropriated metrics associated to reverberation quality.

### 5.1. Mapping $T_{60}^b$ to $T_{hr}$

The only technique that requires a training stage in this work is the proposed algorithm version 2 due to the mapping from blind estimated reverberation time $T_{60}^b$ to threshold $T_{hr}$ required to decide the use of $S_y(k,m)$ at the recovered (dereverberated) speech signal $\tilde{x}(n)$.

Initially a range of thresholds $T = [-16, -15, \ldots, 16]$ dB were tested with each input signal $y(n)$. The most appropriated threshold for each signal was chosen as to maximize the perceptual quality of speech and it was used as a mapping reference.

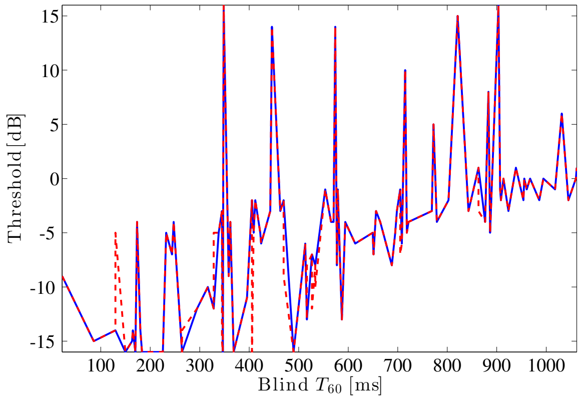

**Fig. 3**. Training PCHIP mapping from $T_{60}^b$ to $T_{hr}$ using database $A_1$. The solid line is the resulting mapping and the dashed line is the reference threshold.

Applying database $A_1$ the PCHIP mapping functions were generated and tested using $A_2$. The Figure 3 presents the results for the training PCHIP mapping using database $A_1$, and Figure 4 presents the obtained mapping functions applied to $A_2$. The root mean squared errors (RMSEs) are 2.51 and 7.86 dB, respectively to $A_1$ and $A_2$.

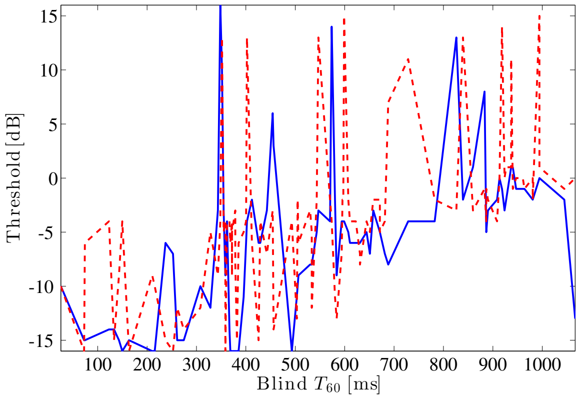

**Fig. 4**. Testing PCHIP mapping from $T_{60}^b$ to $T_{hr}$ using database $A_2$. The solid line is the resulting mapping and the dashed line is the reference threshold.

### 5.2. Comparative analysis

This Section presents the evaluation of the dereverberation techniques under the scope of some reverberation quality measures. The applied methods which use the reference (clean) and degraded speech signals for estimating the quality assessment are the perceptual quality of reverberated speech $Q_m$ [4], reverberation time $T_{60}$ [6], room spectral variance $\sigma_r^2$ [7] and direct-to-reverberant energy ratio $R$ [8, 9].

**Table 1**. Mean performance of the dereverberation algorithms for database $A_1$.

| Quality metric | Unprocessed database | Dereverberation Algorithms | | | |
|---|---|---|---|---|---|
| | | ICS | SS | PI | PII |
| $Q_m$ | 3.4 | 2.6 | 3.6 | 3.6 | 3.8 |
| $T_{60}$ | 526 | 1237 | 379 | 416 | 294 |
| $\sigma_r^2$ | 5.5 | 5.0 | 5.7 | 5.4 | 5.9 |
| $R$ | 7.6 | 1.6 | 7.4 | 7.3 | 8.6 |

Tables 1 and 2 consist of the mentioned reverberation quality methods assessing four dereverberation algorithms, ideal channel selection (ICS) [1], spectral subtraction (SS) [2], proposed algorithms version 1 (PI) and 2 (PII), observed in the scenarios of the training ($A_1$) and test ($A_2$) databases, respectively. Furthermore, the baseline scores were established on the evaluation of the unprocessed database. Although only PII requires a training dataset, all techniques were equally evaluated on all databases for the sake of a fair comparative analysis.

**Table 2**. Mean performance of the dereverberation algorithms for database $A_2$.

| Quality metric | Unprocessed database | Dereverberation Algorithms | | | |
|---|---|---|---|---|---|
| | | ICS | SS | PI | PII |
| $Q_m$ | 3.4 | 2.8 | 3.6 | 3.5 | 3.8 |
| $T_{60}$ | 509 | 1027 | 375 | 411 | 251 |
| $\sigma_r^2$ | 5.7 | 5.1 | 5.8 | 5.5 | 6.1 |
| $R$ | 7.6 | 2.0 | 7.4 | 7.6 | 7.4 |

Consistently for all tables, the measures $Q_m$ and $R$ show a certain amount of increase comparing the proposed approaches with ICS and SS algorithms, and the measures $T_{60}$ and $\sigma_r^2$ present the opposite behavior, i.e., reduction in value. The only technique that does not show any improvement compared to the unprocessed data scores is the ICS algorithm, which despite being designed to address the combined situation of reverberation and noise, it would be expected to work efficiently in just reverberation scenario, as stated in [1]. The ICS weak performance could be explained by the fixed and aggressive threshold $\tau$ for the dataset in analysis [1].

Observing Table 1, the algorithm PII achieved relative improvements in $Q_m$ of about 12%, 46%, 6% and 6% with respect to the unprocessed database, ICS, SS and PI algorithms, respectively. For Table 2 the improvements were about 12%, 36%, 6% and 9%.

The algorithm PI introduces a blind dereverberation technique based on ICS algorithm, which is expected to reduce the consonant confusion errors, as stated in [1]. Although PI algorithm could reach higher performance when compared to unprocessed data and ICS, it could not overcome the SS algorithm performance reaching an equivalent performance for database $A_1$ (Table 1) and a slightly worse performance for database $A_2$ (Table 2), due to the fact that it uses the same fixed threshold $\tau$ for any input signal $y(n)$, motivating and confirming the benefits of using a threshold $T_{hr}$ dependent on the reverberation characteristics of $y(n)$ implemented in PII algorithm.

The algorithm PII has also shown remarkable relative performances concerning the $T_{60}$ for both Tables 1 and 2 reaching about 22% and 33% of improvements, respectively, when compared to SS algorithm, which reached the second best $T_{60}$ perfomances.

## 6. CONCLUSION

This work proposed two blind dereverberation algorithms based on ideal channel selection [1], which originally requires the clean and degraded speech signals to apply the technique. A blind dereverberation algorithm means that it depends only on the degraded speech signal, having no need of the reference signal.

The first proposed technique combined the ICS with the spectral subtraction [2], making the spectral subtracted speech signal replaces the clean speech in the original ICS framework. The second proposed technique was devised using the structure as the first one, except that threshold for selecting the channels is not fixed anymore, and could vary with the algorithm input signal. The first technique introduced the blind approach based on ICS and led the second technique to the implementation of a threshold dependent on the input signal characteristic approach.

The two main contributions of this work were: 1) the modification on ICS algorithm making it dependent only the degraded speech and increasing the dereverberation effectiveness, and 2) the introduction of a rationale to make a reverberation amount variable channel selection threshold.

The effectiveness of the proposed approach PII when compared to the unprocessed data and ICS algorithm reaches approximately 12% and 36% for test database $A_2$.

## 7. REFERENCES

[1] O. Hazratia and P. C. Loizou, "Tackling the Combined Effects of Reverberation and Masking Noise Using Ideal Channel Selection," *Journal of Speech, Lang., and Hearing Research*, vol. 55, pp. 500–510, Apr. 2012.

[2] T. de M. Prego, A. A. de Lima and S. L. Netto, "On the Enhancement of Dereverberation Algorithms Based on a Perceptual Evaluation Criterion," *Proc. InterSpeech*, Lyon, France, Aug. 2013.

[3] M. Wu and D. Wang, "A Two-Stage algorithm for one-microphone reverberant speech enhancement," *IEEE Trans. on Audio, Speech and Lang. Proc.*, vol. 14, May 2006.

[4] A. A. de Lima, T. de M. Prego, S. L. Netto, B. Lee, A. Said, R. W. Schafer, T. Kalker, and M. Fozunbal, "On the quality-assessment of reverberated speech," *Speech Communication*, vol. 54, pp. 393–401, Mar. 2012.

[5] T. de M. Prego, A. A. de Lima, S. L. Netto, B. Lee, A. Said, R. W. Schafer, T. Kalker, and M. Fozunbal, "A blind algorithm for reverberation-time estimation using subband decomposition of speech signals," *Journal of Acoustical Society of America*, vol. 131, no. 4, pp. 2811-2816, Apr. 2012

[6] M. Karjalainen, P. Antsalo, A. Mäkivirta, T. Peltonen, and V. Välimäki, "Estimation of modal decay parameters from noisy reponse measurements," *Proc. Conv. Audio Engineering Society*, Amsterdam, Netherlands, pp. 867–878, May 2001.

[7] J. J. Jetz, "Critical distance measurement of rooms from the sound energy spectral response," *J. Acoustic. Soc. Am.*, vol. 65, pp. 1204–1211, May 1979.

[8] P. Zahorik, "Assessing auditory distance perception using virtual acoustics," *J. Acoustic. Soc. Am.*, vol. 111, pp. 1832–1846, Apr. 2002.

[9] P. Zahorik, "Direct-to-reverberant energy ratio sensitivity," *J. Acoustic. Soc. Am.*, vol. 112, pp. 2110–2117, Nov. 2002.

[10] M. Jeub, M. Schäfer, and P. Vary, "A Binaural Room Impulse Response Database for the Evaluation of Dereverberation Algorithms," *Proc. 16th Int. Conf. on Digital Signal Processing*, Santorini, Greece, 2009.