6th Congress of Alps-Adria
Acoustics Assosiation
16.-17. October 2014
Graz, Austria

Austrian Acoustics Association
Österreichische Gesellschaft für Akustik

Alps Adria Acoustics Association

ISBN: 978-3-200-03921-6

AN IMPROVED ALGORITHM FOR MAXIMUM-LIKELIHOOD-BASED BLIND ESTIMATION OF REVERBERATION TIME

Andrea Andrijašević, Miroslav Vrankic, Juraj Šimunić

Faculty of Engineering, University of Rijeka, Croatia (andrea.andrijasevic@riteh.hr)

Abstract: *Reverberation is a well known phenomenon in the area of room acoustics and speech processing representing the gradual sound energy decay in an enclosure after the sound source has been switched off. Due to detrimental effects of temporal and frequency smearing of the speech signals received on microphones that it causes, numerous methods for signal enhancement have been proposed, where a number of them require the reverberation time (RT) parameter value to be known in advance. In this paper an improved algorithm for blind time-domain estimation of reverberation time from received speech signals is presented, where novel rules for defining the optimal boundaries of the speech segments used for maximum-likelihood estimation are introduced. We examine the performance of our algorithm, more precisely, the deviation of the estimated RT values and their variance from the true value of the parameter, and show that the estimation accuracy of our algorithm over similar methods is improved over a wide range of reverberation time values.*

Key words: reverberation time, blind estimation, speech, maximum-likelihood, coefficient of determination

1. INTRODUCTION

Reverberation is caused by the multi-path propagation of acoustic signals from a source to a receiver in an enclosure, causing temporal and spectral smearing of the original signal at the receiver's side. Due to these detrimental effects on speech signals received in hearing aids, hands-free telecommunication devices or systems for automatic speech recognition, different methods for speech enhancement have been developed. A part of them relies on the most common parameter for quantifying the level of reverberation in an enclosure, namely the reverberation time parameter *RT* [1,2,3]. Reverberation time is defined as the time interval needed for steady-state sound energy to decay by 60 dB once the excitation source has been switched off [4].

Reverberation time estimation has a long history, starting from the 19th century and Sabine's pioneering work, when he made an empirical formula for reverberation time calculation based solely on room geometry and absorption characteristics of surface materials used in the environment [5]. Following that, methods using different excitation signals for reverberation time estimation were created; the interrupted noise method of analyzing energy decay curves obtained by switching off the noise source on the one hand, and methods using pseudorandom maximum length sequences, sine sweeps or noise bursts in combination with Schroeder's method [6] for reverberation time calculation via backwards integration of obtained room impulse responses (RIRs) on the other hand. Nowadays, there is a high demand for algorithms that are able to estimate *RT* accurately from recorded reverberant speech, where an *RT* estimation algorithm can be broadly classified as either a blind one that requires no *a priori* knowledge and estimates the *RT* value directly from the received microphone signal, or as a machine learning-based algorithm that requires the steps of training and validation on large speech and impulse response corpuses prior to its usage.

The group of blind *RT* estimation algorithms was established with the work of Ratnam *et al.* [7, 8] in which they modeled RIRs with Polack's statistical reverberation model for diffuse sound fields [9] and proposed an assumption of speech decay phases and RIR model equality which enabled them to develop a simple log-likelihood function for maximum-likelihoood (ML) estimation of reverberation time values. Many different algorithms based on Ratman's work have been proposed since [10-16], where the changes presented were either about increasing the noise robustness [10], not using the whole input signal for *RT* estimation but only the segments consisting of speech decays [11,13,16], changing RIR model fine structure function distribution

[12], decreasing the non-linear optimization computational load [14], or using the autocorrelation function (AC) of the prewhitened microphone signal as an estimate of the RIR AC [15].

On the other hand, the second group of methods, the machine learning-based ones, try to link the features extracted from the input reverberant signal with the room reverberation time value. They mainly differ in the complexity and type of feature vectors used. Simpler methods, such as in [17-19] try to link the negative side variance of the distribution of reverberant speech signal energy decay rates obtained either in time-frequency domain via least squares [17,18] or as energy ratios of prewhitened speech segments in the time domain [19], to reverberation time parameter value, whilst more complex algorithms use the decay rate histograms as feature vectors [20]. The method presented in [21] estimates the reverberation time using the ratio of signal envelope modulation spectrum energy in higher (>20 Hz) and lower modulation frequencies as input features, whilst the method proposed in [22] uses features obtained by filtering the STFT representation of reverberant signal with 2D Gabor filters. Whilst powerful, machine learning methods inherently posses a large shortcoming when compared to the blind methods of being restricted to functioning properly only in enclosures with acoustic features similar enough to the ones of the RIRs used to train the machine.

In this paper we build on the method developed by Löllmann *et al.* [11] which is considered as the state-of-the-art algorithm with the lowest bias and variance in noise-free conditions [23]. We propose an improved algorithm, where we show that it is possible to decrease the bias and variance even further by estimating two additional parameter values (beside the RT) that enable us to discard most of the spurious estimates that are causing the bias and increasing the variance of the estimator.

The paper is organized as follows: in Chapter 2 we explain the sound decay model used for the ML estimation procedure, briefly present the algorithm proposed in [11], show its limitations and propose modifications for overcoming them. In Chapter 3 we describe the experimental procedure used for testing the performance of the proposed algorithm along with Löllmann's algorithm as baseline estimator, while in Chapter 4 we present and discuss the results obtained. Finally, in Chapter 5 we present our conclusions.

2. ML ESTIMATION METHOD FORMULATION

2.1. Speech decay model

Following the mathematical relation that the observed reverberant speech signal can be seen as a result of convolution of the original speech signal and room impulse response, it is possible to estimate reverberation time value using segments of observed speech consisting of free decays following abrupt speech offsets [10,11].

We model those speech decays as discrete random processes:

$$d(n) = A\, v(n)\, \mathrm{e}^{-\rho\, n\, T_s}, \quad n \geq 0 \tag{1}$$

where n marks the discrete time index, A represents sound amplitude, ρ sound decay rate, T_s marks the sampling period while $v(n)$ is a sequence of independent random variables that are distributed normally with zero mean and variance of one $\mathcal{N}(0,1)$. By modeling the speech decay as a random process of independent but, due to exponential term in (1), non-identically distributed random variables, it is possible to derive the corresponding log-likelihood function for an observed speech decay sequence $d(n)$ of length N in the form of:

$$LL(d{:}\rho) = -\frac{N}{2}\Bigg((N-1)(-T_s\rho) + \ln\left(\frac{2\pi}{N} \sum_{i=0}^{N-1} \mathrm{e}^{2iT_s\rho}\, d^2(i) \right) + 1 \Bigg). \tag{2}$$

As (2) is a function of parameter ρ solely, true decay rate value can be estimated as the value of argument of (2) for which the function obtains its maximal value. The corresponding value of reverberation time can then be obtained by using the following relation:

$$RT = \frac{3\ \ln 10}{\rho}\,. \tag{3}$$

Unfortunately, not all phonemes have sharp offsets, and due to that, speech decays used for ML blind estimation will usually not be free decays but convolutions of phoneme decays with room impulse responses [7,17]. Thus, in general, estimation method will tend to produce estimate values higher than the true room reverberation time for rooms with low RT values.

2.2. Original RT estimator

The original method for RT estimation, as proposed in [11], can briefly be summarised as follows. The received reverberant speech signal is first downsampled by a factor D and segmented into frames of length N_{frame} shifted by N_{shift} samples.

Each frame, designated by the index k, is then divided into L sub-frames of equal length. After that, it is checked whether the energy, maximum and minimum value of the current sub-frame of index l deviates from the successive sub-frame of index l+1 according to

$$\begin{gathered} \mathrm{E}(k,l) > \mathrm{E}(k,l+1), \\ \max(k,l) > \max(k,l+1), \\ \min(k,l) < \min(k,l+1), \end{gathered} \tag{4}$$

where $\mathrm{E}(k,l)$ is the current sub-frame energy, $\max(k,l)$ is the maximum value of the sub-frame, while

$\min(k, l) <$ stands for the minimum value of sub-frame sample.

If any of the conditions in (4) is violated, it is checked whether $l \geq l_{min}$ and if it is not, the comparison is aborted and the next frame is processed. Otherwise, the consecutive sub-frames for which the conditions in (4) are true are combined and detected as a possible sound decay. For this segment, the reverberation time is estimated according to the procedure described in Chapter 2. Finally, the last K_h acquired estimates of reverberation time are kept in a histogram and the location of the maximum of the histogram is taken as the true value of the parameter.

The values of estimates produced by the method proposed in [11], labeled as Löllmann's method in the reminder of the paper, are highly dependent on the values of two parameters used for sound energy decay phase detection, namely sub-frame size N_{frame}/L and minimal number of sub-frames l_{min}. Their product defines the minimal length of sound energy decay needed to start the estimation procedure. In the case the value of that product is set to a value corresponding to a time length of 0.5 s or higher, it will be impossible to detect reverberation time values lower than 0.5 accurately (as it is not possible to measure a 60 dB or higher energy decay in real environments due to background noise levels), and thus an overestimation of low *RT* values will occur [13] with good high *RT* tracking. On the other hand, if the value of the product is set to very low values (lower than 0.1 s), it will be impossible to track high reverberation times accurately. In this paper we decide for the middle ground and use a value of 0.2 s for the minimal duration of sound energy decay used for estimation.

2.3. Proposed modifications

In order to decrease both bias and the variance of the estimates obtained by Löllmann's method comprising the histogram from which the final estimate of room reverberation time value is drawn, we propose two improvements to the original method.

Firstly, for each detected sound decay curve, we propose to identify the indices of the last maximum of the first sub-frame and of the first minimum of the last sub-frame, and to use the sound decay contained solely in between those indices for ML estimation. The reason for this modification can be observed in Figure 1. In this figure, a portion of speech signal that has been detected as a sound decay is presented (blue+red). It can be observed that, due to sub-frame size, a sound segment which does not show decay property has unfortunately been included (blue). By applying the aforementioned criterion, the true sound decay portion can be extracted (red).

Secondly, we propose to calculate two additional values from the decay curve for each *RT* estimate obtained via ML estimation: the corresponding dynamic range of the decay $dBfall$ measured in decibels and the coefficient of determination R^2 that determines how much variance in speech decay curve the model described in (1) can explain. In that way, we hope to discard the spurious estimates obtained by either decay curves of low dynamic range or curves that do not comply well with the model of sound decay that we imposed.

It should be noted that the value of dynamic range should be calculated using the estimated *RT* by the following expression:

$$dBfall = 60 \cdot T_d / RT, \quad dB \tag{5}$$

and not from the speech decay curve itself due to signal fluctuations. In (5), T_d stands for the duration of speech decay sequence expressed in seconds. Finally, in Figure 2 the overall block processing scheme has been presented.

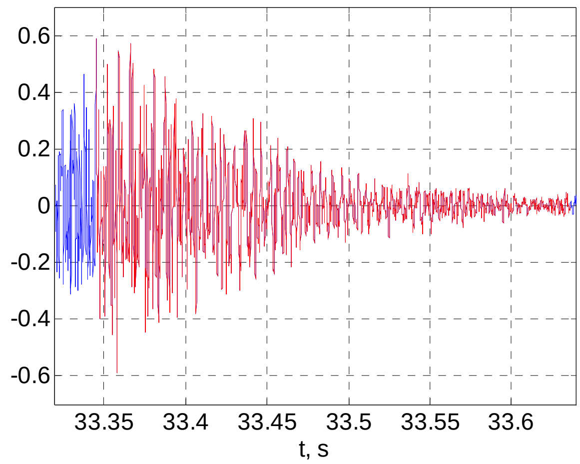

Fig. 1. Example of a speech decay segment: Löllmann's method (blue+red), proposed method (red)

3. EXPERIMENTAL PROCEDURE

For our experiments, 36 clean speech recordings taken from the OLLO (OLdenburg LOgatome) speech corpus [24] were used, where two different texts written in German were uttered by eighteen speakers of both genders. Each of the recordings was approximately one minute long.

The room impulse responses were generated after Polack's statistical reverberation model for reverberation times ranging from 0.2 to 1.2 s in 0.1 s increments. After that, each of the recordings was convolved with each of the generated RIRs, and the estimates of reverberation time were collected for each (recording, RIR) pair separately using Löllmann's original method as implemented in [25] as well as with our own method with the modifications and criterions as described in Chapter 2.

The values of parameters used for *RT* estimation both in Löllmann's algorithm as well as our proposed algorithm are given in Table 1. Before comparing the algorithms' performances in Chapter 4, it should be stated that the eligible *RT* values for the Löllmann's algorithm were defined in discrete time steps of 0.05 s and confined to interval [0.2, 2] s, whilst our algorithm was implemented using MATLAB Optimisation toolbox functions for non-linear constrained optimization and thus the possible *RT* values obtained by ML procedure were more continuous.

l_{min}	5		f_s	16 kHz
L	10		D (Löllmann)	2
$N_{frame} \cdot f_s$	40 ms		K_h	All acquired estimates satisfying the constraints
$N_{shift} \cdot f_s$	20 ms			

Table 1. Values of parameters used for testing the baseline and proposed algorithm

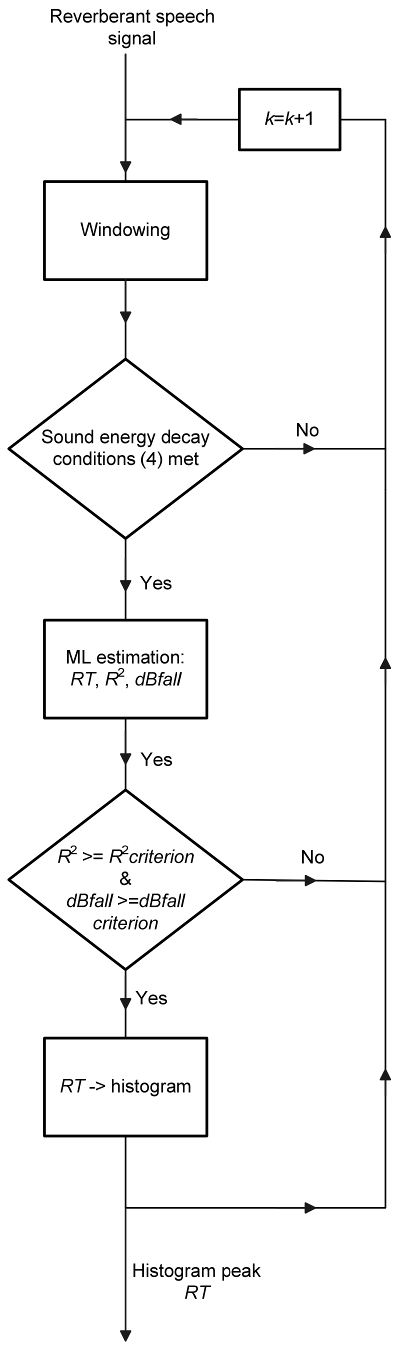

Fig. 2. Proposed method flow chart

4. RESULTS

In Figure 3, the boxplots showing the distributions of estimation errors versus generated RIRs reverberation time true values obtained from all 36 recordings by using baseline Löllmann's algorithm, are presented. The central red marks represent the median values, the lower and upper boxes limits present the 25th and 75th percentiles of the distribution, whiskers represent the most extreme data points and crosses represent outliers. It can be seen that the variance correlated measure (the difference between the 75th and 25th percentile) increases with the increase of the RIR *RT* value. Also, it should be noted that there exists a positive bias (overestimation) for low *RT* values that can be attributed to the gradual offsets of speech sounds, and a negative one (underestimation) for high values of *RT*.

By applying our estimation method that determines the true indices of the start and end of a speech sound decay segment and keeps only those *RT* estimates that satisfy the imposed decibel decrease criterion, with criterion set to 25 dB we obtain the boxplots of estimation errors as presented in Figure 4. The percentile difference has decreased significantly for reverberation times higher than 0.7 s when compared to Löllmann's method with a drawback of a slightly higher and positive estimation bias due to differences in algorithms' implementations.

When both decibel criterion and the criterion of degree of compliance of the decay curve with the model have been utilised, and set to values of 25 dB and 0.6 respectively, the distributions of errors as presented in Figure 5 were gained. It can be observed that the estimation bias for lower *RT* values (<0.5 s) has decreased (when compared to Fig. 3 and Fig. 4) and that the spread of the errors has reduced even more.

In Fig. 6, the boxplots of the estimation errors obtained for each of the texts separately, obtained by using our proposed method, are shown: the distributions of errors are almost identical for the first (T1) and second (T2) text used in all cases.

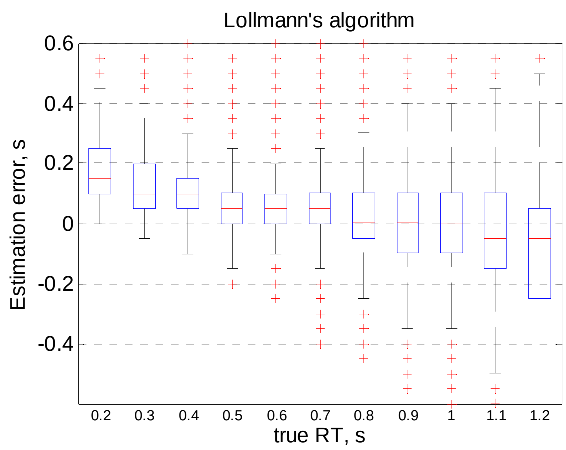

Fig. 3. Estimation errors for Löllmann's method for all 36 speech recordings

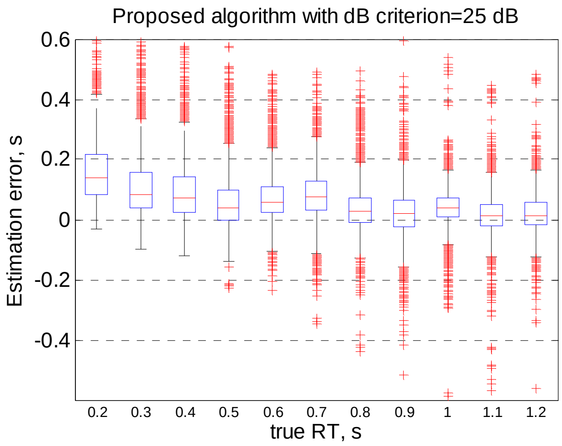

Fig. 4. Estimation errors for proposed method for all 36 speech recordings, with *dBfall criterion*=25 dB and R^2 *criterion*=0

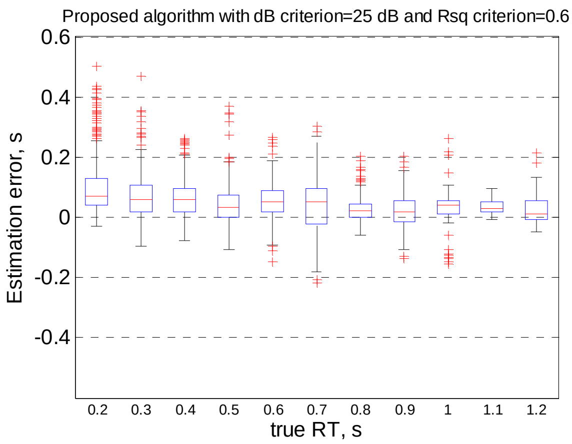

Fig. 5. Estimation errors for proposed method for all 36 speech recordings, with *dBfall criterion*=25 dB and R^2 *criterion*=0.6

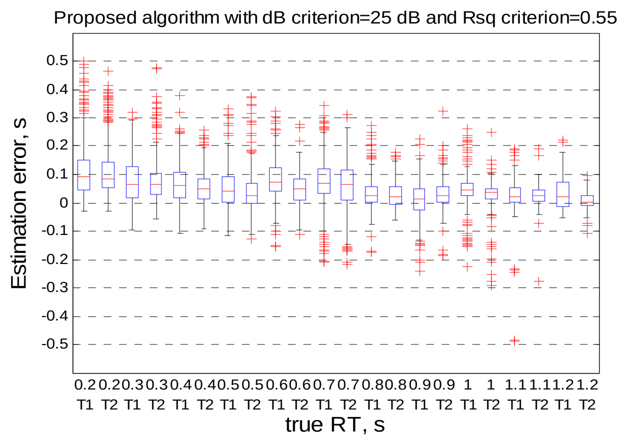

Fig. 6. Estimation errors for the first (T1) and second (T2) text separately, with *dBfall criterion*=25 dB and R^2 *criterion*=0.55

It can be concluded that regardless of the differences in implementation, the improvements in bias and variance reduction obtained with the proposed algorithm are significant and surpass the discretization steps chosen for the baseline estimator.

5. CONCLUSIONS

In this paper, modifications to an existing maximum-likelihood reverberation time estimation algorithm have been presented. It was shown that by estimating the values of speech energy decay sequence dynamic range and the degree of its compliance to the model for the sound decay imposed, and using them to discriminate between the true and spurious *RT* estimates, both the bias and the variance of the estimator were significantly reduced.

The improved algorithm for ML estimation has been tested on simulated impulse responses based on the statistical model of reverberation for diffuse enclosures only. Thus, the next logical step in the investigation will be examination of the degree of bias and variance reduction obtained with proposed criterions when real measured impulse responses are utilised.

REFERENCES

[1] P. A. Naylor, and N. D. Gaubitch, Eds: ***Speech Dereverberation***, Springer, 2010

[2] E. A. P. Habets: ***Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement***, Ph.D. thesis, Technische Universiteit Eindhoven, 2007

[3] H. W. Löllmann, and P. Vary: **Low delay noise reduction and dereverberation for hearing aids**, *EURASIP Journal on Advances in Signal Processing*, Volume 2009, pp. 1-9

[4] H. Kuttruff: ***Room Acoustics***, Elsevier Science Publishers Ltd., London, 3rd ed., 1991

[5] F. Alton Everest: ***Master handbook of acoustics***, McGraw-Hill, 2009

[6] M. R. Schroeder: **New method of measuring reverberation time**, *J. Acoust. Soc. Am.* 1965, pp. 409–412

[7] R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. OBrien, C. R. Lansing, and A. S. Feng: **Blind estimation of reverberation time**, *J. Acoust. Soc. Am.*, vol. 114, Nov. 2003, pp. 2877–2892

[8] R. Ratnam, D. L. Jones and W. D. OBrien: **Fast algorithms for blind estimation of reverberation time**, *IEEE Signal Process. Letters*, vol. 11, Jun. 2004, pp. 537–540

[9] J. D. Polack, ***La Transmission de l'´energie sonore dans les salles***, Ph.D. thesis, Universit´e du Maine, Le Mans, France, 1988

[10] H. W. Löllmann, and P. Vary: **Estimation of the Reverberation Time in Noisy Environments**, *Proc. Int. Workshop Acoust. Echo and Noise Control (IWAENC)*, Washington, USA, Sep. 2008, ID 9033

[11] H. W. Löllmann, E. Yilmaz, M. Jeub, and P. Vary: **An improved algorithm for blind reverberation time estimation**, *Proc. Int. Workshop Acoust. Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, Aug. 2010, pp. 1-4
[12] T. Jan and W. Wang: **Blind reverberation time estimation based on Laplace distribution**, *Proc. 20th European Signal Processing Conference* (EUSIPCO 2012), Bucharest, Romania, 2012, pp. 2050-2054
[13] P. Kendrick, ***Blind estimation of room acoustic parameters from speech and music signals***, Ph.D. thesis, University of Salford, UK, 2009
[14] C. Schuldt, P. Handel: **Blind low-complexity estimation of reverberation time**, IEEE Workshop on *Applications of Signal Processing to Audio and Acoustics* (WASPAA), USA, New York, October 2013, pp. 1-4
[15] A. Keshavarz, S. Mosayyebpuor, M. Biguesh, T. A. Gulliver, M. Esmaeili: **Speech-model based accurate blind reverberation time estimation using an LPC filter**, *IEEE Transactions on audio, speech, and language processing*, vol. 20, no. 6, August 2012, pp. 1884-1893
[16] T. M. Prego, A. A. Lima, S. L. Neto: **A blind algorithm for reverberation-time estimation using subband decomposition of speech signals**, *J. Acoust. Soc. Am.*, vol. 131, April 2012, pp. 2811–2816
[17] J. Y. C. Wen, E. A. P. Habets, and P. A. Naylor: **Blind estimation of reverberation time based on the distribution of signal decay rates**, *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.*, 2008, pp. 329–332
[18] J. Eaton, N. D. Gaubitch, and P. A. Naylor: **Noise-robust reverberation time estimation using spectral decay distributions with reduced computational cost**, *Proc. International Conference on Acoustics, Speech and Signal Processing* (ICASSP), Vancouver, Canada, 2013, pp. 161-165
[19] N. Lopez, Y. Grenier, G. Richard, and I. Bourmeyster: **Low variance blind estimation of the reverberation time**, *Proc. International Workshop on Acoustic Signal Enhancement* (IWAENC), Aachen, Germany, 2012, pp. 1-4
[20] R. Talmon, E. A. P. Habets: **Blind reverberation time estimation by intrinsic modeling of reverberant speech**, *Proc. International Conference on Acoustics, Speech and Signal Processing* (ICASSP), Vancouver, Canada, 2013, pp. 156-160
[21] T. H. Falk, W. Y. Chan: **Temporal dynamics for blind measurement of room acoustical parameters**, *IEEE Transactions on instrumentation and measurement*, 2010, vol. 59, no. 4, pp. 978-989
[22] F. Xiong, S. Goetze, and B. T. Meyer: **Blind estimation of reverberation time based on spectro-temporal modulation filtering**, *Proc. International Conference on Acoustics, Speech and Signal Processing* (ICASSP), Vancouver, Canada, 2013, pp. 443-447
[23] N. D. Gaubitch, H. Löllmann, M. Jeub, T. Falk, P. A. Naylor, P. Vary, and M. Brookes: **Performance Comparison of Algorithms for Blind Reverberation Time Estimation from Speech**, *Proc. Intl. Workshop Acoust. Signal Enhancement* (IWAENC), Aachen, Germany, 2012, pp. 1-4
[24] OLLO Logatome Corpus: http://medi.uni-oldenburg.de/ollo/
[25] Blind Reverberation Time Estimation code: http://www.mathworks.com/matlabcentral/fileexchange/35740-blind-reverberation-time-estimation