



**HAL**  
open science

## Blind estimation of room acoustic parameters using kernel regression

Arthur Belhomme, Yves Grenier, Roland Badeau, Eric Humbert

► **To cite this version:**

Arthur Belhomme, Yves Grenier, Roland Badeau, Eric Humbert. Blind estimation of room acoustic parameters using kernel regression. AES 60th Conference, Feb 2016, Leuven, Belgium. hal-01248010

**HAL Id: hal-01248010**

**<https://hal.science/hal-01248010v1>**

Submitted on 21 Mar 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Blind estimation of room acoustic parameters using kernel regression

Arthur Belhomme<sup>1,2</sup>, Yves Grenier<sup>1</sup>, Roland Badeau<sup>1</sup> and Eric Humbert<sup>2</sup>

<sup>1</sup>*Institut Mines-Télécom; Télécom ParisTech; CNRS LTCI, Paris, 75013, France*

<sup>2</sup>*Invoxia, Issy-les-Moulineaux, 92130, France*

Correspondence should be addressed to Arthur Belhomme ([arthur.belhomme@telecom-paristech.fr](mailto:arthur.belhomme@telecom-paristech.fr))

## ABSTRACT

Room acoustic parameters are key information for dereverberation or speech recognition. Usually, when one needs to assess the level of reverberation, only the reverberation time  $RT_{60}$  or a direct to reverberant sounds index  $D_\tau$  is estimated. Yet, methods which blindly estimate the reverberation time from reverberant recorded speech do not always differentiate the  $RT_{60}$  from the  $D_\tau$  to evaluate the level of reverberation. That is why we propose a method to jointly blindly estimate these parameters, from the signal energy decay rate distribution, by means of kernel regression. Evaluation is carried out with real and simulated room impulse responses to generate noise-free reverberant speech signals. The results show this new method outperforms baseline approaches in our evaluation.

## 1. INTRODUCTION

When a sound is emitted in an enclosed space, the microphone does not only capture the output of the source: all the paths the sound may follow, from the source to the microphone, are added to the direct one and produce reverberation. These paths depend on the different reflections of the acoustic wave on the walls and surfaces it meets, that can be viewed by means of the room impulse response (RIR). This transfer function of the room would be the record of an impulse played from a spherical source and depends both on the room characteristics and the source-microphone distance.

Many parameters can be extracted from the RIR, however the reverberation time ( $RT_{60}$ ) is the most common to describe the reverberation of the room. That is why one can find a large amount of methods in the literature to compute it from a measured RIR ([1], [2]) or estimate it blindly from a recording in a room. Among them, a major stream uses a maximum-likelihood procedure to estimate the time-constant of the signal energy decay, directly related to the  $RT_{60}$  ([3], [4], [5], [6], [7]). Other methods use a machine learning approach by training deep neural networks in order to map the spectrum envelopes of reverberant signals to the corresponding room acoustic parameters ([8], [9], [10]). Finally, a bench of methods use the statistical distribution of the decay rate

of the energy envelope, in order to link some statistical moments with the reverberation time ([11], [12], [13], [14]); this stream is the most similar to our approach.

However the reverberation time is specific to the characteristics of the room, its dimensions and material, but independent of the source-receiver distance. Yet this distance has a great impact on the amount of reverberation [15], characterized by the ratio of direct to reverberant sounds. Even if the direct-to-reverberant ratio (DRR) is the most straightforward index to assess the level of reverberation [16] many variants of this measure exist, including the clarity and definition indexes, well-suited for syllable and music intelligibility [17]. Thus, for a given room (*i.e.* for a given  $RT_{60}$ ) one can obtain a certain range of clarity or definition index, according to the distance between the source and the microphone. Thereby, methods such as [12] and [11] which aim to link the  $RT_{60}$  to a single feature of the energy decay rate distribution, through a fixed  $n^{\text{th}}$ -order polynomial, will not distinguish a low  $RT_{60}$  and a high source-microphone distance from a high  $RT_{60}$  and a short source-microphone distance. To fully characterize reverberation, one thus needs to estimate the  $RT_{60}$  and an index related to the DRR.

On that account, we propose a method to jointly estimate the reverberation time and a clarity (or definition) index, from the energy decay rate distribution of a re-

reverberant signal. We use kernel regression, specifically the *Nadaraya-Watson* estimator [18], on a database of energy rate distributions obtained from a wide range of synthetic and real RIRs. The paper is organised as follows: Section 2 presents the room acoustic parameters to be estimated and the used features. In Section 3 we derive the estimator and present our method. The evaluation database is detailed in Section 4 and the results are compared to state of the art methods ([19],[15]) in Section 5. Finally, in Section 6 some conclusions are drawn.

## 2. FEATURE EXTRACTION AND ROOM ACOUSTIC PARAMETERS

### 2.1. Room acoustic parameters

As claimed in Section 1, our goal is to jointly estimate the reverberation time and an index of the reverberation level. We define in this section the different expressions of these parameters.

#### Reverberation time

Defined as the time interval to measure an energy decay of 60 dB, the  $RT_{60}$  can be computed with the Schroeder backward integration method [2] via the Energy Decay Curve (EDC):

$$EDC(n) = \sum_{k=n}^{N_h} h(k)^2, \forall n \geq 0 \quad (1)$$

where  $h$  is the room impulse response of length  $N_h$ . Then, the  $RT_{60}$  is the required time for the EDC to decrease by 60 dB.

#### Level of reverberation

Two close indexes to assess the balance between the direct and reverberant sounds are the clarity index, defined in [16]

$$C_\tau = 10 \log_{10} \left( \frac{\sum_{n=0}^{N_\tau} h^2(n)}{\sum_{n=N_\tau+1}^{\infty} h^2(n)} \right) \text{dB}, \quad (2)$$

and the definition index, defined in [16]

$$D_\tau = 10 \log_{10} \left( \frac{\sum_{n=0}^{N_\tau} h^2(n)}{\sum_{n=0}^{\infty} h^2(n)} \right) \text{dB}, \quad (3)$$

where  $N_\tau$  is the number of samples corresponding to the time  $\tau$  ranging from 0.1 ms to 1 s, and  $h(n)$  is the room

impulse response. Two widely adopted values of  $\tau$  are 50 ms or 80 ms, since they correspond to the duration splitting the (useful) early reflections to the (disturbing) late reverberation [20]. As explained in [17], the  $D_{50}$  is an objective criterion to measure the speech intelligibility, while  $C_{80}$  is more designed for music transparency. However Parada *et al.* showed in [15] that the  $C_{50}$  is more correlated with the Perceptual Evaluation of Speech Quality (PESQ) and the phoneme accuracy rate than the  $D_{50}$ . Even if the performance of our method is quite the same for the  $C_{50}$  and  $D_{50}$ , we focus on the  $C_{50}$  in order to later compare our results, as there is no other method for  $D_{50}$  estimation to our knowledge.

### 2.2. Feature extraction

#### 2.2.1. Decay rate distribution

The statistical reverberation model developed by Polack in [21] is often used to describe RIRs in a diffuse field. The RIR is represented as one realization of a non-stationary stochastic process, a Gaussian white noise of variance  $\sigma^2$  damped by a decreasing exponential envelope, linked to a room parameter  $\delta$ :

$$h(n) = b(n)e^{-\delta \frac{n}{f_e}} \quad (4)$$

with  $b(n) \sim \mathcal{N}(0, \sigma^2)$ ,  $\delta = \frac{3 \ln 10}{RT_{60}}$  and  $f_e$  the sampling rate. This model has been generalized by Habets to the entire RIR in [20], using a different noise (Gaussian with another variance) for the direct path. The energy envelope of the RIR can be expressed as:

$$e(n) = \mathbb{E}[h(n)^2] = \sigma^2 e^{-2\delta \frac{n}{f_e}} = \sigma^2 e^{\lambda_h \frac{n}{f_e}} \quad (5)$$

with  $\mathbb{E}[\cdot]$  the expectation operator and  $\lambda_h = -2\delta$  the energy decay of the room.

However, we aim to analyze a reverberant speech, which can be viewed as the convolution of a RIR and an anechoic speech. Wen *et al.* derived in [11] an expression of the energy envelope of a reverberant signal  $d_x(n)$ , after a speech endpoint according to the energy decay rates of the anechoic signal and the room,  $\lambda_s$  and  $\lambda_h$  respectively:

$$d_x(n) = \begin{cases} (e^{\lambda_h n} - e^{\lambda_s n}) / (\lambda_h - \lambda_s) & \text{if } \lambda_h \neq \lambda_s \\ ne^{\lambda_h n} & \text{if } \lambda_h = \lambda_s \end{cases} \quad (6)$$

As the sum of two exponential terms will be dominated by the exponential term with the largest value  $\lambda$ , the en-

ergy decay rate of the reverberant signal  $\lambda_x$  can be approximated as:

$$\lambda_x \approx \max[\lambda_h, \lambda_s] \quad (7)$$

If we process a short term analysis of the signal, with frames of  $N_\omega$  samples and a hop size of  $R$  samples, we can compute the total energy over the  $m^{\text{th}}$  frame as  $E_m = \sum_{i=0}^{N_\omega-1} d_x(mR+i)$  and then estimate the decay rate by computing the logarithmic ratio between two successive frames:

$$\rho(m) = \log\left(\frac{E_m}{E_{m-1}}\right). \quad (8)$$

Thereby, the energy decay rates distribution in equation (8) provides information about the decay rate of the room and thus the  $RT_{60}$ . We now see how to use the distribution to build an estimator of the  $(RT_{60}, C_{50})$  or  $(RT_{60}, D_{50})$  pair.

### 2.2.2. Characteristic function

The methods developed in [11] and [12] use the negative-side variance of the distribution (*i.e.* the variance where the negative values of  $\rho$  have been symmetrized) to estimate the  $RT_{60}$ , with a polynomial model built on a training dataset. In [13], the authors prefer to use the temporal differential of cepstral coefficients and collect the variance, skewness, kurtosis, median absolute deviation of the obtained distribution as input features for support vector regression.

Therefore, the relevance of some statistical moments of the reverberant energy rate distribution, for room acoustic parameters estimation, is clear. However the choice of the statistical moments to be used is unclear, nor shared between the authors. This is why we choose to use the characteristic function of the energy decay rate distribution, as it conveys the global information of the distribution, and indirectly contains all the statistical ordinary moments. As presented in [22], the characteristic function of a real random variable  $X$  is defined as:

$$\phi_X(f) = \int e^{ifx} dF_X(x) = \mathbb{E}[e^{ifX}] \quad (9)$$

where  $F_X(x)$  is the cumulative distribution of  $X$  and  $f$  is the angular frequency. As the characteristic function behaves simply under shift, scale changes and summation of independent variables, it is a convenient tool for parameters estimation [22]. Moreover, the different statistical ordinary moments, related to the room acoustic

parameters, can be extracted from this function. If a random variable has ordinary moments up to the  $k^{\text{th}}$  order, the characteristic function has a  $k^{\text{th}}$  derivative at the zero frequency and:

$$\mathbb{E}[X^k] = (-i)^k \phi_X^{(k)}(0) \quad (10)$$

where  $\phi_X^{(k)}$  is the  $k^{\text{th}}$  derivative of  $\phi_X$ . This is why we decide to represent the distribution of the energy decay rate of reverberant speech by its characteristic function. It will constitute the observation feature in a kernel regression approach.

## 3. BLIND ESTIMATION OF ACOUSTIC ROOM PARAMETERS

### 3.1. The Nadaraya-Watson estimator

Consider a random input vector  $X$  and a random output vector  $Y$ , with joint probability density function (JPDF)  $P_{X,Y}(x,y)$ . We seek a function  $f$  which best predicts  $Y$  given  $X$ , by minimizing the expected prediction error (EPE)  $\mathbb{E}[(Y - f(X))^2]$ . The solution is given by the regression function [18]:

$$f(x) = \mathbb{E}[Y|X=x] = \int y P_{X,Y}(y|x) dy. \quad (11)$$

A way to estimate this regression function is to use a kernel to approximate the JPDF. As defined in [23], a kernel  $K$  is a similarity measure of the form

$$K : \begin{cases} \mathcal{X} \times \mathcal{X} & \longrightarrow \mathbb{R} \\ (x, x') & \longmapsto \langle \phi(x), \phi(x') \rangle \end{cases}$$

where  $\phi$  is a mapping function from the input space  $\mathcal{X}$  to a dot product space  $\mathcal{V}$ :

$$\phi : \begin{cases} \mathcal{X} & \longrightarrow \mathcal{V} \\ x & \longmapsto \phi(x) \end{cases}$$

In regression applications, kernels are parameterized by a constant  $\lambda$  that dictates the width of the neighborhood to be considered, and are denoted  $K_\lambda$ . With  $N$  realizations  $(x_i, y_i)$  of the random variables  $(X, Y)$ , and a kernel  $K_\lambda$ , one can estimate the JPDF  $P_{X,Y}(x,y)$  [18] as:

$$\hat{P}_{X,Y}(x,y) = \frac{1}{N} \sum_{i=1}^N K_\lambda(x, x_i) K_\lambda(y, y_i). \quad (12)$$

Let us recall (11), noticing  $P_{X,Y}(y|x) = P_{X,Y}(x,y)/P_X(x)$  and  $P_X(x) = \int P_{X,Y}(x,y)dy$ , we obtain:

$$f(x) = \frac{\int y P_{X,Y}(x,y) dy}{\int P_{X,Y}(x,y) dy}. \quad (13)$$

If we use the estimator (12) we can approximate the regression function by:

$$\begin{aligned} \hat{f}(x) &= \frac{\int y \frac{1}{N} \sum_{i=1}^N K_\lambda(x, x_i) K_\lambda(y, y_i) dy}{\int \frac{1}{N} \sum_{i=1}^N K_\lambda(x, x_i) K_\lambda(y, y_i) dy} \\ &= \frac{\sum_{i=1}^N K_\lambda(x, x_i) \int y K_\lambda(y, y_i) dy}{\sum_{i=1}^N K_\lambda(x, x_i) \int K_\lambda(y, y_i) dy} \end{aligned} \quad (14)$$

With the properties of kernel functions proposed in [23],  $\int y K_\lambda(y, y_i) dy = y_i$  and  $\int K_\lambda(y, y_i) dy = 1$ , which leads to the *Nadaraya-Watson* estimator:

$$\hat{f}(x) = \frac{\sum_{i=1}^N y_i K_\lambda(x, x_i)}{\sum_{i=1}^N K_\lambda(x, x_i)}. \quad (15)$$

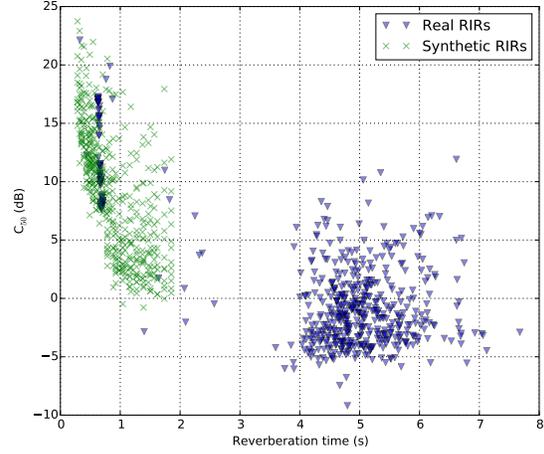
### 3.2. Features and output vectors

Thus, our method needs a dataset of  $N$  pairs  $(x_i, y_i) \in \mathbb{R}^p \times \mathbb{R}^2$  where each  $x_i$  is the characteristic function  $\phi_{\rho_i}(f)$  of energy decay rate distribution of a reverberant speech, and  $y_i$  is the corresponding room acoustic parameters:  $(RT_{60}, C_{50})$  or  $(RT_{60}, D_{50})$ . Actually, the characteristic function is computed for  $\frac{p}{2}$  frequencies and stored in a vector of  $\mathbb{R}^p$  where the first  $\frac{p}{2}$  components are the real parts of  $\phi_{\rho_i}(f)$  and the last  $\frac{p}{2}$  components are their imaginary parts. The range of the frequencies and their sampling are chosen such that  $|\phi_{\rho_i}(f)|$  lies between almost 0.1 and 1.

When a reverberant signal is recorded, we compute its energy decay rate distribution and the corresponding characteristic function, which is the feature vector. The output vector  $\hat{y}$ , which carries the room acoustic parameters, is estimated with the *Nadaraya-Watson* estimator (15).

## 4. PERFORMANCE EVALUATION

To evaluate our method we use anechoic speech signals of 25 english speakers (12 males, 12 females, one



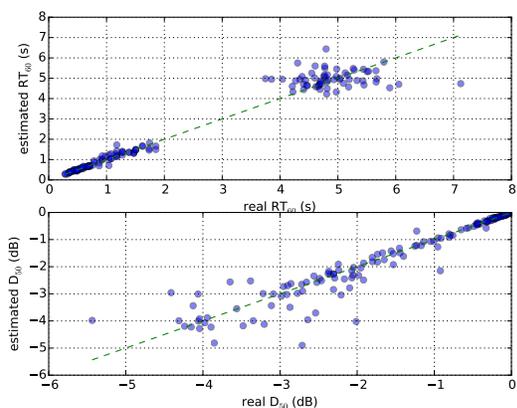
**Fig. 1:** Room distribution in the  $(RT_{60}, C_{50})$  plane

child; around 5 min each) from the TSP Speech database [24]. Our RIR database is composed of 1015 responses; among them 455 synthetic responses generated from the Fast Image-Source Method [25], with the  $RT_{60}$  ranging from 0.1 s to 2.0 s and the source-microphone distance ranging from 0.1 m to 10 m. We add 560 real RIRs, from different open-access databases (Aachen Impulse Response [26], MARDY [27], OpenAIR [28], Queen-Mary [29]) with reverberation times ranging from 0.3 s to 8 s and  $C_{50}$  from -10 dB to 25 dB. They can be observed in the  $(RT_{60}, C_{50})$  plane in Figure 1.

Reverberant speech signals are obtained by convolving each RIR with each anechoic speech and the energy decay rates are computed using frames of 32 ms with 50 % overlap. The obtained distributions are centered and reduced, then the characteristic functions are computed for angular frequencies from 0 to 0.4 with 0.001 increments. Finally, we compute the average over the 25 different speakers, for each RIR.

We test our method in a 7-fold approach; we randomly split our corpus of characteristic functions in 7 subcorpora (fold) and successively estimate the room acoustic parameters of a fold, using the remaining 6 as dataset. The prediction error is then the average of errors obtained over the 7 tests. We use a Gaussian kernel with a bandwidth of  $\lambda = 5 \cdot 10^{-4}$ :

$$K_\lambda(x, x_i) = \frac{1}{\lambda} e^{-\frac{\|x-x_i\|^2}{2\lambda}}. \quad (16)$$



**Fig. 2:** Joint estimation of  $RT_{60}$  and  $D_{50}$ , for a random fold

Schroeder's method (described in Section 2.1) is used to compute the reference  $RT_{60}$ ; the reference reverberation level indexes are directly computed from the RIR, using the definitions in Section 2.1.

## 5. RESULTS

We compute two kinds of errors, the relative error  $e_{rel}$  (%) and the root mean square error  $e_{rms}$  (same unit as the data):

$$e_{rel} = 100 * \frac{1}{N} \sum_{i=1}^N \left| \frac{x_i - \hat{x}_i}{x_i} \right|$$

$$e_{rms} = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{N}}$$

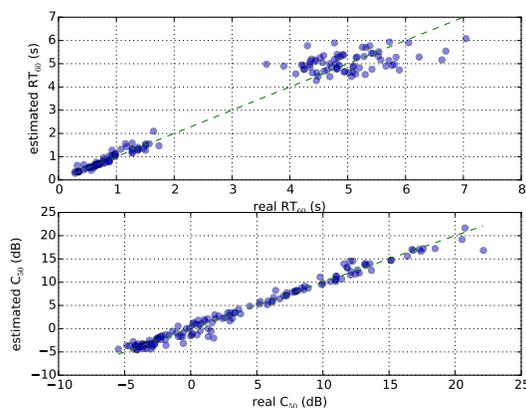
with  $\hat{x}_i$  the estimation of  $x_i$  in a fold of  $N$  tests. However, we will not use the relative errors when dealing with the  $C_{50}$  since some are equals to zero.

### 5.1. Synthetic and real RIRs

We tested our method on a corpus made of both synthetic and real RIRs. The errors are then averaged over the 7 folds and presented in Table 1.

Figures 2 and 3 show the deviation between true and estimated parameters, for two different folds. The relative errors are around 10 % and 15 %, on a corpus including high reverberation times (up to 8 s).

However, the other methods found in the literature are tested for  $RT_{60}$  up to 1 s or 2 s. Then, if we reduce our



**Fig. 3:** Joint estimation of  $RT_{60}$  and  $C_{50}$ , for a random fold

corpus to RIRs whose reverberation time does not exceed 2 s (usual reverberant environments) we obtain better results, summarized in Table 2.

	$RT_{60} \leq 2s$		$RT_{60} \leq 1s$	
	$e_{rms}$	$e_{rel}$	$e_{rms}$	$e_{rel}$
$RT_{60}$	180ms	9.3%	75ms	7.8%
$D_{50}$	0.11dB	9.8%	0.07dB	11.8%
$C_{50}$	0.97dB		0.90dB	

**Table 2:** Estimation error on the low  $RT_{60}$  corpus, synthetic and real RIRs

With these lower reverberant environments estimation errors are lower, less than 10 % relative error for  $RT_{60}$  and  $D_{50}$ , less than 1 dB of root mean square error for  $C_{50}$  (which varies between 0 dB and 25 dB). Then we can compare our results to the performance of state of the art methods [19].

### 5.2. Comparison

We compare our results for  $RT_{60}$  estimation to the ones provided in [19], which tests three methods of the literature (Modulation Energy Ratio (MER), Maximum Likelihood (ML), Spectral Decay Distributions (SDD)) on a database made of simulated and real RIRs. Even if we use the same simulation method for synthetic RIRs, and include the same real RIR database (AIR) in our corpus, we do not use the same corpus and then this comparison is intended to give an idea of the relative

	All RIRs		Synth.		Real	
	$e_{rms}$	$e_{rel}$	$e_{rms}$	$e_{rel}$	$e_{rms}$	$e_{rel}$
$RT_{60}$	496 ms	10.4 %	143 ms	9.8 %	716 ms	10.9 %
$D_{50}$	0.42 dB	14.3 %	0.09 dB	9.4 %	0.55 dB	17.1 %
$C_{50}$	1.07 dB		0.73 dB		2.17 dB	

Table 1: Estimation errors

	MER [13]	ML [7]	SDD [11]	Our method
$e_{rel}$	20 %	15 %	10 %	7.8 %

Table 3: Comparison for  $RT_{60}$  prediction

	Baseline [8]	NIRA [15]	Our method
$e_{rms}$	9.05 dB	5.52 dB	4.81 dB

Table 4: Comparison for  $C_{50}$  prediction

performance scores. The reverberation times in [19] vary up to 1 s every 100 ms, and the relative errors for noise-free reverberant speech are approximately 20 %, 15 % and 10 % for [13], [7], [11] respectively. In our 7-fold approach we obtain a mean of 7.8 % relative error, which is close to the magnitude of [11] as we can see in Table 3.

We found in [15] a way of comparing the  $C_{50}$  values. In this paper, the authors test their method (called NIRA) on the real RIRs database MARDY [27] and give the corresponding root mean square error. They also confront their results to the one obtained with their implementation of [8], used as a baseline. We then set a training corpus with all the characteristic functions obtained from our simulated and real RIRs, excluding the ones corresponding to the MARDY database. Then, we estimate the  $C_{50}$  of the MARDY database with our method and this training dataset. For noise-free reverberant speech, the baseline obtains 9.05 dB RMS error, NIRA obtains 5.52 dB and our method 4.81 dB as we can see in Table 4.

## 6. CONCLUSION

As reverberation is determined by static properties of the room (resulting in the reverberation time) and the source-microphone distance (resulting in the clarity or definition index), we have presented a method to blindly jointly estimate these parameters. The distribution of the en-

ergy decay rates of a reverberant speech, described by its characteristic function, serves to perform kernel regression on a training dataset, with a Gaussian kernel and a low bandwidth. We tested our method in a 7-fold approach, with simulated and real room impulse responses and obtained a mean of 10.4 % relative error for  $RT_{60}$  estimation, 14.3 % for  $D_{50}$  estimation, 1.07 dB root mean square error for  $C_{50}$  estimation. Compared to the available results from the literature, our method slightly outperforms the reverberation time estimation. There is no other method to blindly estimate the  $D_{50}$  so we only compared the  $C_{50}$  estimator to two state of the art methods, that were outperformed.

Future work will focus on the estimation of these parameters in noisy environments and introduce sparsity constraints in the feature vector, selecting the best frequencies of the characteristic function to be used.

## 7. REFERENCES

- [1] W.C. Sabine. *Collected papers on acoustics (originally 1921)*, 1993.
- [2] M.R. Schroeder. New Method of Measuring Reverberation Time. *The Journal of the Acoustical Society of America*, 37(3):409, 1965.
- [3] R. Ratnam, D.L. Jones, B.C. Wheeler, C.R. Lansing, and A.S. Feng. Blind estimation of reverberation time. *Journal of the Acoustical Society of America*, 114(5):2877–2892, 2003.
- [4] K Lebart and JM Boucher. A new method based on spectral subtraction for speech dereverberation. *ACUSTICA*, 87(3):359–366, 2001.
- [5] P. Kendrick, F.F. Li, T.J. Cox, Y. Zhang, and J.A. Chambers. Blind estimation of reverberation parameters for non-diffuse rooms. *Acta Acustica united with Acustica*, 93(5):760–770, 2007.

- [6] Y. Zhang, J.A. Chambers, F.F. Li, P. Kendrick, and T.J. Cox. Blind estimation of reverberation time in occupied rooms. In *Signal Processing Conference, 2006 14th European*, pages 1–5, Sept 2006.
- [7] M. Jeub H. Lollmann, E. Yilmaz and P. Vary. An improved algorithm for blind reverberation time estimation. *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, (5):1–4, 2010.
- [8] P. Kendrick, T. J. Cox, F.F. Li, Y. Zhang, and J. A. Chambers. Monaural room acoustic parameters from music and speech. *The Journal of the Acoustical Society of America*, 124(1):278–287, 2008.
- [9] F.F. Li and T.J. Cox. Speech transmission index from running speech: A neural network approach. *The Journal of the Acoustical Society of America*, 113(4):1999–2008, 2003.
- [10] T.J. Cox, F. Li, and P. Darlington. Extracting room reverberation time from speech using artificial neural networks. *J. Audio Eng. Soc.*, 49(4):219–230, 2001.
- [11] 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. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 329–332, March 2008.
- [12] N. Lopez, Y. Grenier, and I. Bourmeyster. Low variance blind estimation of the reverberation time. In *Acoustic Signal Enhancement; Proceedings of IWAENC 2012; International Workshop on*, pages 1–4, Sept 2012.
- [13] T.H. Falk and W.Y. Chan. Temporal dynamics for blind measurement of room acoustical parameters. *Instrumentation and Measurement, IEEE Transactions on*, 59(4):978–989, April 2010.
- [14] J. Eaton, N.D. Gaubitch, and P.A. Naylor. Noise-robust reverberation time estimation using spectral decay distributions with reduced computational cost. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 161–165, May 2013.
- [15] P.P. Parada, D. Sharma, and P.A. Naylor. Non-intrusive estimation of the level of reverberation in speech. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 4718–4722, May 2014.
- [16] P. A. Naylor and N. D. Gaubitch. *Speech Dereverberation*. Springer, eds edition, 2010.
- [17] H. Kuttruff. *Room acoustics / Heinrich Kuttruff*. Elsevier Applied Science London ; New York, 5th ed. edition, 2009.
- [18] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.
- [19] N. D. Gaubitch, H. W. Loellmann, M. Jeub, T. H. Falk, P. A. Naylor, P. Vary, and M. Brookes. Performance comparison of algorithms for blind reverberation time estimation from speech. In *Acoustic Signal Enhancement; Proceedings of IWAENC 2012; International Workshop on*, pages 1–4, Sept 2012.
- [20] E.A.P. Habets. *Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement*. PhD thesis, Technische Universiteit Eindhoven, 2007.
- [21] J.D. Polack. *La transmission de l'énergie sonore dans les salles*. PhD thesis, 1988. Thèse de doctorat dirigée par Bruneau, Michel Physique Le Mans 1988.
- [22] A. Feuerverger and R.A. Mureika. The empirical characteristic function and its applications. *Ann. Statist.*, 5(1):88–97, 01 1977.
- [23] B. Scholkopf and A.J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001.
- [24] McGill University. <http://www-mmmsp.ece.mcgill.ca/documents../data/index.html>. 2002.
- [25] E.A. Lehmann and A.M. Johansson. Diffuse reverberation model for efficient image-source simulation of room impulse responses. *Audio, Speech, and Language Processing, IEEE Transactions on*, 18(6):1429–1439, Aug 2010.

- [26] M. Jeub, M. Schäfer, and P. Vary. A binaural room impulse response database for the evaluation of dereverberation algorithms. In *Proceedings of the 16th International Conference on Digital Signal Processing, DSP'09*, pages 550–554, Piscataway, NJ, USA, 2009. IEEE Press.
- [27] J.Y.C. Wen, N.D. Gaubitch, E.A.P. Habets, T.Myatt, and P.A. Naylor. Evaluation of speech dereverberation algorithms using the mardy database, Sept. 2006.
- [28] <http://www.openairlib.net/auralizationdb>.
- [29] R. Stewart and M. Sandler. Database of omnidirectional and b-format room impulse responses. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 165–168, March 2010.