

# Blind reverberation time estimation from ambisonic recordings

Andrés Pérez-López<sup>\*†</sup>, Archontis Politis<sup>‡</sup>, and Emilia Gómez<sup>\*§</sup>

<sup>\*</sup> Department of Information and Communication Technologies, Music Technology Group, Universitat Pompeu Fabra  
Carrer Tanger 122-144, 08005 Barcelona, Spain

<sup>†</sup> Multimedia Technologies Unit, Eurecat, Centre Tecnològic de Catalunya  
Carrer Bilbao 72, 08005 Barcelona, Spain

<sup>‡</sup> Faculty of Information Technology and Communication Sciences, Audio Research Group, Tampere University  
Korkeakoulunkatu 1, 33720 Tampere, Finland

<sup>§</sup> Centre for Advanced Studies, Joint Research Centre, European Commission  
Calle Inca Garcilaso, 3, 41092 Seville, Spain

**Abstract**—Reverberation time is an important room acoustic parameter, useful for many acoustic signal processing applications. Most of the existing work on blind reverberation time estimation focuses on the single-channel case. However, the recent developments and interest on immersive audio have brought to the market a number of spherical microphone arrays, together with the usage of ambisonics as a standard spatial audio convention. This work presents a novel blind reverberation time estimation method, which specifically targets ambisonic recordings, a field that remained unexplored to the best of our knowledge. Experimental validation on a synthetic reverberant dataset shows that the proposed algorithm outperforms state-of-the-art methods under most evaluation criteria in low noise conditions.

**Index Terms**—blind reverberation time estimation, ambisonics, dereverberation, acoustic parameter estimation

## I. INTRODUCTION

Knowledge about the acoustic properties of an enclosure is a fundamental topic with many applications in the microphone array and acoustic signal processing field. Problems such as dereverberation [1] or source separation [2] may benefit from this information, and may require prior estimation of the related parameters. Reverberation time  $T_{60}$  [3] might be one of the most widespread acoustic parameters; it represents the time required for the reverberant sound field power to decay by 60 dB. Reverberation time can be accurately computed from the room geometry [4] or from the impulse response (IR) [5]; the problem of  $T_{60}$  estimation just from observations of the reverberant signal itself is referred to as the blind reverberation time estimation, and it still remains an open research question.

The 2015 Acoustic Characterisation of Environments (ACE) Challenge [6] gathered dozens of methods designed for blind  $T_{60}$  and direct-to-reverberation ratio (DRR) estimation; nowadays, it is still considered as a state-of-the-art source for performance evaluation and comparison among methods.

Most of the model-based  $T_{60}$  estimation algorithms consider the reverberant signal envelope as an exponential decay, so that the problem is reduced to finding a signal offset and estimate the decay rate. Moreover, in last years, data-driven models

have outperformed the previous state-of-the-art results [7]–[9]. A comparative review on single-channel blind  $T_{60}$  estimation algorithms was recently published [10].

However, most of the existing reverberation time estimation methods focus on the single-channel case. A representative example can be drawn from the ACE Challenge, where, despite the fact that one of the reverberant datasets was recorded with an *em32 Eigenmike* spherical microphone array, none of the methods use of it for the  $T_{60}$  estimation task.

On the other hand, recent years have witnessed a growing interest in immersive audio for virtual and augmented reality. This situation has consolidated Ambisonics [11] as the *de facto* standard for spatial audio. Dedicated spherical microphone arrays have reached the market in last years; their multichannel nature makes possible spatial manipulations that complement traditional signal enhancement methods.

In this paper, we present a novel approach to the problem of multichannel blind reverberation time estimation, specifically focusing on first order ambisonic (FOA) recordings. The method is based on a dereverberation stage followed by system identification. To the best of our knowledge, the proposed algorithm is the first reverberation time estimation method specifically designed for first order ambisonic audio<sup>1</sup>.

## II. SIGNAL MODEL

Let us consider a FOA signal  $x_m(t)$ , with  $m = 0 \dots M-1$  as the channel number, and  $M = 4$ . Let us further assume that  $x_m(t)$  represents the signal captured by an ideal spherical microphone array located in a reverberant enclosure, where a static sound source  $s(t)$  is present. The ambisonic room impulse response between source and receiver is represented by  $h_m(t)$ . In the absence of noise, the recorded signal is the convolutive mix of the source signal  $s(t)$  and the IR:

$$x_m(t) = h_m(t) * s(t) \quad (1)$$

<sup>1</sup>Full implementation is available under an open-source license at [https://github.com/andresperezlopez/ambisonic\\_rt\\_estimation](https://github.com/andresperezlopez/ambisonic_rt_estimation).

Here,  $T_{60}$  estimation assumes no receiver directionality. Therefore, in what follows, all relevant parameters will be estimated from the zeroth order ambisonic channel,  $x_0(t)$ .

### III. BASELINE METHOD

The baseline algorithm, taken from [12], is based on the detection of abrupt event offsets in the time-frequency domain. The subband energy decay on the transitions can be then used to compute an estimate of the full-band decay. This method performed best in the ACE Challenge regarding the Pearson correlation coefficient between estimated and true  $T_{60}$  [6].

Let us consider the zeroth order channel of the recorded signal,  $x_0(t)$ , and its short-time Fourier transform (STFT) counterpart  $X_0(k, n)$ , where  $k$  and  $n$  indicate frequency bin and time frame indices respectively. The *subband energy*  $\bar{E}(k, n)$  of the recorded signal can be expressed as:

$$\bar{E}(k, n) = |X_0(k, n)|^2. \quad (2)$$

A free decay region (FDR) is defined as a group of consecutive bins within the same subband which exhibit a monotonically decreasing energy. A FDR search is performed on the subband energy spectrogram  $\bar{E}(k, n)$ : for each band, the algorithm tries to find at least one FDR, iteratively reducing the FDR length if no candidates are found.

The next step is the estimation of the reverberation time, which is performed using a subband equivalent of Schroeder's method [5]. The subband energy decay function (SEDF) associated with a given FDR is computed as:

$$\bar{c}(k, n) = 10 \log_{10} \frac{\sum_{\nu=n}^{L_c-1} \bar{E}(k, \nu)}{\sum_{\nu=0}^{L_c-1} \bar{E}(k, \nu)} \text{dB}, \quad (3)$$

where  $n = 0 \dots, L_c - 1$  spans the length of the FDR. A linear regression is then performed on each SEDF curve:  $T_{60}$  is computed as the time required by the resulting line to reach the  $-60$  dB reference.

This procedure yields a  $T_{60}$  estimate per FDR. In order to obtain a global estimate, the algorithm proposes a two-step statistical filtering. First, it obtains a narrowband estimate as the median of all estimates within each subband. Then, the resulting broadband value  $\bar{T}_{60}$  is computed as the median of all subband estimates. The last step of the method is the expansion of the resulting dynamic range by a linear mapping. This procedure is required because of the compression introduced by the median operator. The final value  $T_{60}$  is thus a linear mapping of  $\bar{T}_{60}$ , where the parameters  $\alpha$  and  $\beta$  might be obtained by linear regression on a training stage:

$$T_{60} = \alpha \bar{T}_{60} + \beta \quad (4)$$

### IV. PROPOSED METHOD

We propose a novel method for reverberation time estimation, based on two steps: signal dereverberation, and system identification. The main idea consist in obtaining an estimate of the dereverberated signal, which is later used for estimating the multichannel IR given the recorded reverberant signal. The reverberation time can be thus computed by the decay slope of the estimated IR.

#### A. Dereverberation

Let us consider the convolutive transfer function (CTF) model of the signal model in Eq. 1 in the STFT domain:

$$X_m(k, n) = \sum_{l=0}^{L_h-1} H_m(k, l) S(k, n-l), \quad (5)$$

where the multichannel filter  $H_m(k, l)$  of length  $L_h$  contains the CTF coefficients between the source and the microphones.

It is possible to split the former expression into two consecutive elements, which would conceptually match the early and late parts of the room's impulse response:

$$X_m(k, n) = D_m(k, n) + R_m(k, n) = \sum_{l=0}^{\tau-1} H_m(k, l) S(k, n-l) + \sum_{l=\tau}^{L_h-1} H_m(k, l) S(k, n-l), \quad (6)$$

where the parameter  $\tau$  represents the mixing time, which states the transition time between early reflections and late reverberation. In other words, the captured signal is split between a *direct* part  $D_m(k, n)$ , containing the direct path and the early reflections, and a *reverberant* part  $R_m(k, n)$ , which mainly contains the diffuse part of the reverberation.

Assuming a multichannel auto-regressive (MAR) model,  $R_m(k, n)$  can be expressed as a multichannel infinite impulse response (IIR) filter applied to the recorded signal:

$$R_m(k, n) = \sum_{i=1}^M \sum_{l=0}^{L_g-1} X_i(k, n-\tau-l) G_{mi}(k, l), \quad (7)$$

where the coefficients  $G_{mi}(k, l) \in \mathbb{C}$  model the relation between channels  $m$  and  $i$ , and have a length of  $L_g$  frames.

By grouping all time frames  $n = 1 \dots, N-1$ , it is possible to express Eq. 7 in vector notation:

$$\mathbf{R}_m(k) = \tilde{\mathbf{X}}_\tau(k) \mathbf{G}_m(k), \quad (8a)$$

$$\tilde{\mathbf{X}}_\tau(k) = [\tilde{\mathbf{X}}_{\tau,1}(k), \dots, \tilde{\mathbf{X}}_{\tau,M}(k)], \quad (8b)$$

where  $\tilde{\mathbf{X}}_{\tau,m}(k)$  is a  $N \times L_g$  matrix, and  $\mathbf{R}_m(k)$  and  $\mathbf{G}_m(k)$  are column vectors with lengths  $N$  and  $L_g M$ , respectively.

Finally, the expression can be further simplified by omitting the frequency dependence, and by expressing the channels as columns in the vector notation. Substituting this expression in Eq. 6 leads to the MAR equation:

$$\mathbf{D} = \mathbf{X} - \tilde{\mathbf{X}}_\tau \mathbf{G}. \quad (9)$$

Here, the dereverberation problem consists in the estimation of the MIMO filter  $\mathbf{G}$ , so that the *clean* signal  $\mathbf{D}$  (containing both direct path and early reflections) can be computed.

The solution proposed in this paper is based on the method described in [13]. In this case, the dereverberation problem is tackled as an optimization problem, considering that the spectrograms of the reverberant signal are less sparse than those of the corresponding *clean*, and ensuring that the inter-channel signal properties are maintained. Although the presented method is applied on the whole signal in *batch* mode, alternative *online* methods could be also used, e.g. [14].

By using iteratively reweighted least squares (IRSL) [15], it can be shown that an iterative solution for the estimation of  $\mathbf{G}$  at the iteration  $(i)$  is given by the following expression:

$$\mathbf{G}^{(i)} = (\tilde{\mathbf{X}}_\tau^H \mathbf{W}^{(i)} \tilde{\mathbf{X}}_\tau)^{-1} \tilde{\mathbf{X}}_\tau^H \mathbf{W}^{(i)} \mathbf{X}, \quad (10)$$

where  $\mathbf{W}^{(i)}$  is a  $N \times N$  diagonal matrix whose diagonal values,  $w_n^{(i)}$ , can be updated as:

$$w_n^{(i)} = (\mathbf{d}_n^{H(i-1)} \Phi^{-1(i-1)} \mathbf{d}_n^{(i-1)})^{\frac{p-2}{2}} + \epsilon. \quad (11)$$

In turn,  $\mathbf{d}_n$  represents the rows of  $\mathbf{D}$  arranged as column vectors of length  $M$ ,  $\Phi$  is the  $M \times M$  spatial covariance matrix (SCM) of  $\mathbf{D}$ ,  $\epsilon$  is an arbitrary small positive value, and  $p \leq 1$ . The computation and update of the SCM matrix is given by:

$$\Phi^{(i)} = \frac{1}{N} \mathbf{D}^{T(i)} \mathbf{W}^{(i)} \mathbf{D}^{*(i)}. \quad (12)$$

To conclude the dereverberation method, Eqs. 9, 10, 11 and 12 can be applied iteratively, starting by updating Eq. 11, until convergence is reached:

$$\|\mathbf{D}^{(i)} - \mathbf{D}^{(i-1)}\|_F / \|\mathbf{D}^{(i)}\|_F < \eta, \quad (13)$$

where  $\eta$  is an arbitrary small positive value, or alternatively until the maximum number of iterations  $i_{max}$  is exceeded. For the initialization, the following values are proposed:  $\mathbf{D} = \mathbf{X}$  and  $\Phi = \mathbf{I}_M$  (the identity matrix of size  $M \times M$ ).

### B. System Identification

The output of the dereverberation step is the multichannel signal  $\mathbf{D}_m$ , which ideally contains the direct plus early reflection components of the source. Therefore, given the reverberant signal  $\mathbf{X}_m$  and the dereverberated signal  $\mathbf{D}_m$ , an estimate of the late room impulse response might be derived by identifying the filter connecting the two. As stated in Section II, we are primarily interested on the response of the omnidirectional channel; for that reason, the filter estimation is performed with the zeroth order components of both recorded and dereverberated signals. We perform system identification directly in the STFT through a linear fit between input and output independently for every frequency bin:

$$\hat{H}_0(k) = \frac{\mathbf{d}_0^H(k) \mathbf{x}_0(k)}{\mathbf{d}_0^H(k) \mathbf{d}_0(k)}, \quad (14)$$

where  $\mathbf{d}_0, \mathbf{x}_0$  are  $N \times 1$  length vectors. To avoid complex cross-band modeling of the system response, we use a long STFT window, assumed longer than twice the length of the IR so that a reduction of the CTF to a multiplicative transfer function (MTF) holds [16].

As a last step, the estimated time-frequency filter  $\hat{H}_0(k, n)$  is transformed into the time domain filter  $\hat{h}(t)$ . The  $T_{60}$  is then computed by linear fitting of the Schroeder integral in the  $[-5, -15]$  dB range ( $T_{10}$  estimation method), after filtering  $\hat{h}(t)$  with an octave-band filter centered at 1 kHz.

TABLE I  
BASELINE SYSTEM: LINEAR REGRESSION PARAMETERS

Dataset	$\alpha$	$\beta$	$\sigma$
Speech	6.6619	-1.4517	0.2131
Drums	8.2421	-2.1939	1.0055

## V. EXPERIMENTAL SETUP

### A. Dataset

The proposed method is evaluated using two different reverberant datasets, containing recordings of *speech* and *drums* respectively. In order to have full control over the reverberation conditions in the experimental setup, the audio clips under consideration have been rendered by the convolutive mixture of clean monophonic recordings with FOA IRs.

The *speech* dataset is composed of the LibriSpeech [17] *test-clean* audio samples longer than 25 s, making a total of 30 audio clips. It contains English language sentences by male and female speakers, often with a small level of background noise. We have used only a 20 s long excerpt of each clip, preceded by an initial offset of 5 s. The *drums* dataset is the *test* subset of the isolated drum recordings from the DSD100 dataset [18]. It contains 50 different audio clips, covering a wide range of music and mixing styles. The same audio lengths and offsets as in the previous case are applied.

The IRs are FOA room impulse responses simulated by the image method with the *Multichannel Acoustic Signal Processing* library [19]. There are 9 different IRs of 1 s, with random  $T_{60}$  values in the range between 0.4 s and 1.1 s approximately, estimated by the  $T_{10}$  method at the 1 kHz band. The angular position of the sources is randomized for each IR, while the receiver position is fixed at the room center, which has a size of  $10.2 \times 7.1 \times 3.2$  m. The source distance is set to half the *critical distance*, thus providing positive DRRs.

The combination of the dry audio clips with the IRs yields a total of 270 and 450 audio clips for the *speech* and *drums* datasets, respectively, after removing the audio clips which mostly contain silence. Those datasets will be referred in the following as the *evaluation* datasets.

Finally, the baseline method requires a previous *fitting* step for the computation of the mapping parameters  $\alpha$  and  $\beta$  from Eq. 4. The procedure has been performed as follows. For the *speech* dataset, we selected again the subset of audio clips longer than 25 s, but in this case on the *dev-clean* dataset, which yields a total of 20 audio clips. For the *drums* dataset, we used the 50 clips of the *development* subset. The generation of the convolutive mixes has followed the same procedure as in the previous case. We will refer to the resulting datasets as the *development* datasets.

### B. Setup

The sampling frequency for all methods is 8 kHz. For the baseline system, the window size is 1024 samples long, with an overlap of 256 samples. The FDR length is set to 500 ms, which has been reported as the ideal theoretical minimum [12]; it corresponds to a FDR length of  $L_c = 15$  samples. At any

TABLE II  
EXPERIMENT RESULTS

Metric	speech		drums	
	Baseline	MAR+SID	Baseline	MAR+SID
Bias	-0.0599	<b>0.0305</b>	<b>0.1521</b>	0.2568
MSE	0.6366	<b>0.0594</b>	<b>13.9376</b>	16.5261
$\rho$	0.8212	<b>0.9848</b>	0.3705	<b>0.7552</b>

frequency band, the value of  $L_c$  is iteratively decreased if no FDR is found, until a minimum value of 3 samples (96 ms). If still no FDR is found, the sound clip is discarded.

In order to compute  $\alpha$  and  $\beta$ , we run the baseline method on both *development* datasets. For each IR, the mean and standard deviation of the results are computed across all sound clips. Then, these values are used for a *weighted least squares* linear regression against the true  $T_{60}$  values. The results are shown in Table I, where  $\sigma$  represents the joint standard deviation of  $\alpha$  and  $\beta$  after the linear regression; the resulting values are in the same range as the values reported in [12].

In the dereverberation stage, the STFT uses a small window size of 128 samples, with 64 samples overlap. The value of  $p$  is set to 0.25, given the good results reported in [13]. Other parameter values are  $\tau = 2$ ,  $i_{max} = 10$ ,  $\eta = 10^{-4}$  and  $\epsilon = 10^{-4}$ . After an exploratory search, the length of the IIR filter  $L_g = 20$  has been chosen as a compromise between method performance and computation time. We have observed a tendency towards poor dereverberation and non-convergence of the IRSI when using small values of  $L_g$  and short audios.

For the SID, the recorded and dereverberated signals are reshaped into much larger STFTs, with a window size of 8 s and a hop size of 0.5 s. The predicted filter size is 1 s.

For both *evaluation* datasets, the two presented methods are employed; we will refer to them as *Baseline* and *MAR+SID*. Furthermore, with the aim of evaluating the performance of the SID method in an isolated manner, we have included a third method, *Oracle SID*. As its name suggests, it performs the System Identification step using the true anechoic signal.

### C. Evaluation metrics

We have considered the three metrics from the ACE Challenge [6], all of them based on the difference between estimated and true values: the *bias*, or mean error; the Mean Squared Error (*MSE*); and the Pearson correlation coefficient. The evaluation has been performed after discarding the outliers, defined as the reverberation time estimates greater than 1.5 s.

## VI. RESULTS

Figure 1(a) shows the experiment result specified for all audio clips individually. Each boxplot represents the statistics of the mean estimation error (*bias*) for a single audio clip subject to all 9 different IRs. The results are organized by method (rows) and dataset (columns). Figure 1(b) aggregates all experiment results into the same plot, showing the statistical distribution of the *bias* per method and dataset. In this case, the *Oracle SID* results are omitted for clarity. The evaluation metrics for all methods are shown in Table II.

According to the results, the proposed method clearly outperforms the baseline in the *speech* dataset by a tenfold MSE improvement. For the *drums* dataset, our method only outperforms the baseline regarding correlation. Nevertheless, an inspection of the statistical distribution of mean estimation errors in Figure 1(b) brings in an interesting observation: the variability of the results given by our method is substantially smaller than the results of the baseline system. This behaviour is consistent across datasets: the mean error distributions with the *speech* dataset are approximately five times narrower than with the *drums* dataset, regardless of the method.

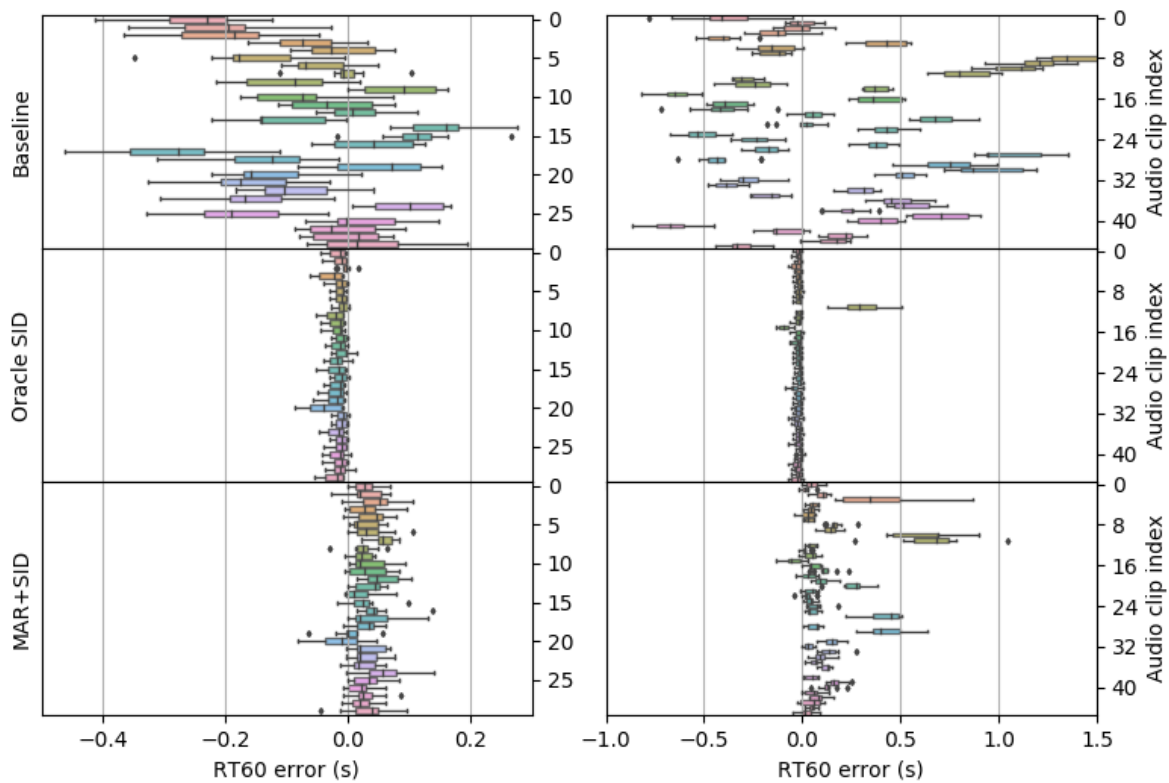
Moreover, all methods behave significantly better on the *speech* dataset. The main reason might be the heterogeneity of the *drums* dataset with respect to dynamic range or timbre, and the potential application of audio effects of any kind. Furthermore, some audio clips of the *drums* dataset contain sounds with a high degree of self-similarity, such as cymbal rolls or exaggerated reverbs; these characteristics would explain the outliers on the proposed method results. It is also interesting to notice the robustness of the proposed method against noise, present in the *speech* dataset. Such robustness is consistent with the behavior reported in [13].

The performance of the *ORACLE SID* method is close to ideal. The *bias* is in all cases under 0.05 s (excepting a *drums* clip containing mostly silence). This result validates the system identification, and allows, in practical terms, a direct evaluation of the proposed method against the groundtruth values.

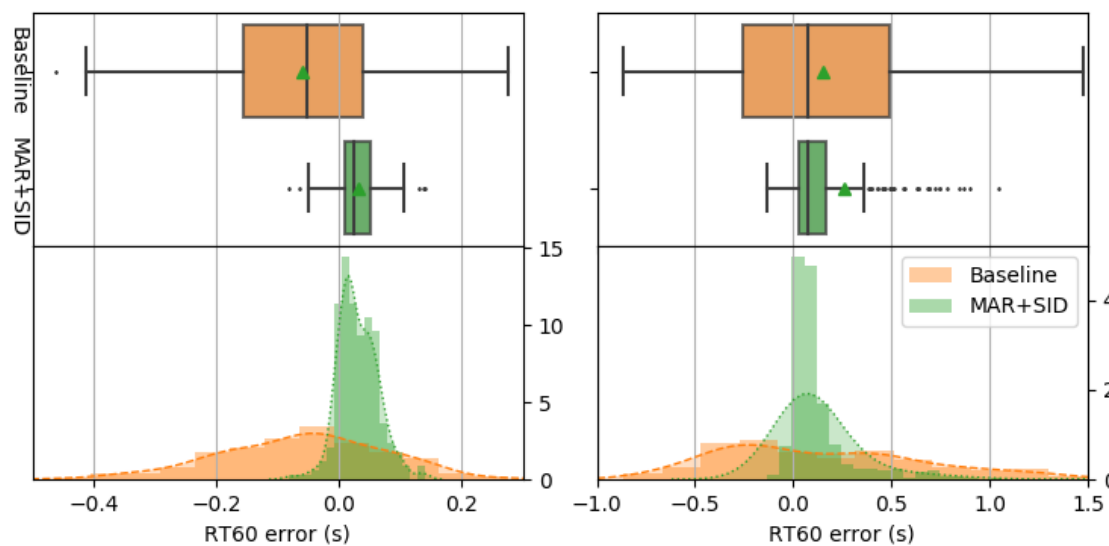
The results obtained in our analysis are very similar to the results reported in recent deep-learning state-of-the-art proposals, e.g. [7]. Such results are not directly comparable for a number of reasons, including the single-channel nature of existing methods, and the different noise ratios under consideration. However, given the similar results obtained with the same evaluation metrics, it might be anticipated that our method may perform as well as other recent data-driven algorithms, under low noise conditions.

## VII. CONCLUSION

We have presented in this work a novel method for blind reverberation time estimation for multichannel audio, with the aim of applying it to the context of ambisonic recordings. Our method is based on a first dereverberation step, performed by a multichannel autoregressive model of the late reverberation. The resulting dry signal is then used to estimate the impulse response decay by means of system identification. The performance of the method is evaluated in a simulated experimental environment with two different reverberant datasets, and compared against a state-of-the-art method. Results show that our method outperforms the baseline method in a majority of evaluation metrics and conditions, and consistently provides results with less variability than the baseline method. In future work, we plan to extend the experimental setup by using recorded IRs. Furthermore, the proposed method could be extended to the case of moving sources by using an *online* autoregressive model. Finally, an extension of the method for higher ambisonic orders remains to be done.



(a) Estimation error computed for each audio clip.



(b) Total estimation error across audio clips and acoustic conditions. Top: boxplot. Bottom: histogram and density plot.

Fig. 1. Experiment results for *speech* (left column) and *drums* (right column) datasets.

## REFERENCES

- [1] S. Braun, A. Kuklański, O. Schwartz, O. Thiergart, E. A. Habets, S. Gannot, S. Doclo, and J. Jensen, "Evaluation and comparison of late reverberation power spectral density estimators," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 6, pp. 1056–1071, 2018.
- [2] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 4, pp. 692–730, 2017.
- [3] H. Kuttruff, *Room acoustics*. Crc Press, 2016.
- [4] W. C. Sabine, *Collected papers on acoustics*. Harvard University Press Cambridge, MA, 1927.
- [5] M. R. Schroeder, "New method of measuring reverberation time," *The Journal of the Acoustical Society of America*, vol. 37, no. 6, pp. 1187–1188, 1965.
- [6] J. Eaton, N. D. Gaubitch, A. H. Moore, and P. A. Naylor, "Estimation of room acoustic parameters: The ace challenge," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 10, pp. 1681–1693, 2016.
- [7] H. Gamper and I. J. Tashev, "Blind reverberation time estimation using a convolutional neural network," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2018, pp. 136–140.
- [8] D. Looney and N. D. Gaubitch, "Joint estimation of acoustic parameters from single-microphone speech observations," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 431–435.
- [9] N. J. Bryan, "Impulse response data augmentation and deep neural networks for blind room acoustic parameter estimation," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1–5.
- [10] H. W. Löllmann, A. Brendel, and W. Kellermann, "Comparative study for single-channel algorithms for blind reverberation time estimation," in *Proc. Intl. Congress on Acoustics (ICA)*, 2019.
- [11] F. Zotter and M. Frank, *Ambisonics*. Springer, 2019.
- [12] T. d. M. Prego, A. A. de Lima, S. L. Netto, B. Lee, A. Said, R. W. Schafer, and T. Kalker, "A blind algorithm for reverberation-time estimation using subband decomposition of speech signals," *The Journal of the Acoustical Society of America*, vol. 131, no. 4, pp. 2811–2816, 2012.
- [13] A. Jukić, T. van Waterschoot, T. Gerkmann, and S. Doclo, "Group sparsity for mimo speech dereverberation," in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2015, pp. 1–5.
- [14] S. Braun and E. A. Habets, "Online dereverberation for dynamic scenarios using a kalman filter with an autoregressive model," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1741–1745, 2016.
- [15] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 3869–3872.
- [16] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time fourier transform domain," *IEEE Signal Processing Letters*, vol. 14, no. 5, pp. 337–340, 2007.
- [17] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: an asr corpus based on public domain audio books," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 5206–5210.
- [18] A. Liutkus, F.-R. Stöter, Z. Rafii, D. Kitamura, B. Rivet, N. Ito, N. Ono, and J. Fontecave, "The 2016 signal separation evaluation campaign," in *Latent Variable Analysis and Signal Separation - 12th International Conference, LVA/ICA 2015, Liberec, Czech Republic, August 25-28, 2015, Proceedings*, P. Tichavský, M. Babaie-Zadeh, O. J. Michel, and N. Thirion-Moreau, Eds. Cham: Springer International Publishing, 2017, pp. 323–332.
- [19] A. Pérez-López and A. Politis, "A python library for multichannel acoustic signal processing," in *AES Virtual Vienna Convention*. Audio Engineering Society, 2020.