INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Effect of Microphone Position Measurement Error on RIR and its Impact on Speech Intelligibility and Quality

Aditya Raikar[1], *Karan Nathwani*[2], *Ashish Panda*[1], *Sunil Kumar Kopparapu*[1]

[1]TCS Research and Innovation – Mumbai, Tata Consultancy Services Limited
[2]Indian Institute of Technology, Jammu

{aditya.raikar,ashish.panda,sunilkumar.kopparapu}@tcs.com, karan.nathwani@iitjammu.ac.in

Abstract

Room Impulse Response (RIR) measurement is done using a microphone and loudspeaker pair and is prone to error due to error in measurement of the microphone position. In literature, the adverse impact of ambient noise on RIR measurement is mostly explored. However, the impact of microphone position error on full, early and late RIR measurements have never been explored. In this paper, we investigate the error in RIR introduced due to error in measurement of the microphone position. We also study the impact of this on the quality and intelligibility of speech. Our analysis shows that the impact of error in microphone position measurement on RIR is as adverse as that of the ambient noise.

Index Terms: Room Impulse Response, Microphone Position Error, Intelligibility and Quality

1. Introduction

Hands free speech communication in an open enclosure creates multi-path reflections before reaching a microphone and the resultant reverberated speech has an impact on the performance of the distant automatic speech recognition engines [1]. The acoustic properties of an enclosure is defined by the RIR or RT60 [2–4] and is an important attribute to achieve enhanced speech quality and intelligibility [1, 5]. RIR measurement is useful in sound reproduction of geometry aware room [6], room geometry reconstruction [7, 8] and robust Automatic Speech Recognition (ASR) [9, 10].

Even when the exact room geometry is known, it is not easy to estimate the transfer function between source and microphone position [11], as RIR estimates are based on the exact location of the microphones and speakers. In modern work environments, the location of microphones get disturbed during routine maintenance or due to human interventions [12, 13] and as a result the RIR is impacted. In such cases, the RIR estimated will be even more distorted. The impact of changed microphone location on RIR results in degradation of sound quality and intelligibility. This is because the prior RIR estimates, based on the original location of the microphones will not correspond to the current (changed) microphone position resulting in a significant mismatch in overall room acoustics [14].

Several studies report RIR estimation under the assumption of *a priori* knowledge about the array of microphone locations (for example [15–19]). Cues such as inter intensity difference (IID) and inter time difference (ITD) which are available with multiple sensors [19], have been used to improve the SNR of the distorted signal using beam-forming (spatial filtering) technique [12,20]. The calibration of an array of microphones in real time is not a trivial task. It is prone to errors in measurement and is a time consuming and tedious exercise [13, 21]. Measurement errors result in degradation of both quality and intelligibility of a speech as shown in [12]. An analytical model of location errors using delay and sum beam-forming is studied in [12] .

The unavailability of acoustic cues in a single microphone scenario, unlike in a microphone array setting, makes modeling of microphone position error and its impact on RIR very challenging. However, some attempts have been made in speech quality improvement using estimated RIR's, when microphone position is assumed to be known *a priori*. In [22], authors have tried to utilize maximum a posteriori (MAP) estimation of RIR in single channel scenario. In another study, MAP has been utilized for joint de-reverberation and de-noising by considering noisy estimate of RIR to be known [23]. In [24], the inverse filtering of RIR has been carried out by maximizing the skewness of LP residual. Yet another study on inverse filtering for RIR estimation [25] utilizes LP residual cepstrum for de-reverberation application.

The adverse impact of ambient noise on RIR measurement has been explored in literature. To the best of our knowledge there has been no study on the impact on full, early and late RIR measurement caused by microphone position error. In this paper, we investigate the effect of RIR variation due to change in microphone position on both quality and intelligibility of speech. Specifically we analyze the variation in RIR due to a variation in the microphone position. Our detailed analysis shows that the impact on RIR due to microphone position error is as adverse as the one caused due to ambient noise. The rest of the paper is as follows. Section 2 introduces the problem statement and captures the disturbance in RIR due to measurement error in position of the microphone. We briefly discuss the traditional RIR estimation techniques used in literature. Impact of Microphone position error on speech quality and intelligibility is analysed in Section 3. This is followed by a brief conclusion and future scope in Section 4.

2. Problem Statement

Modeling RIR is a complex task and one of the most popular method to model it is using the Image method [26]. Utilities to simulate the RIR [27] are available for research purposes [28]. The RIR h is a function of c (the speed of sound), f_s (the sampling frequency), L (volume of the room), β (the reverberation time), s and r (the rectangular coordinates of the source and receiver positions expressed as $[x_s y_s z_s]$). Mathematically, it can be expressed as

$$h = \texttt{rir\_gen}(c, f_s, r, s, L, \beta, n). \tag{1}$$

Note that while all the other parameters that are required to compute h are not prone to error, the measurement of the position of the source and receiver can be erroneous. For the sake

http://dx.doi.org/10.21437/Interspeech.2020-1578

of ease let us assume that there is an error $\epsilon = [\epsilon_x, \epsilon_y, \epsilon_z]$ (introduced because of maintenance of the panel on which the microphone is mounted) in the measurement of the position of the microphone. So the erroneous location of the source is $r_\epsilon = [x_r + \epsilon_x, y_r + \epsilon_y, z_r + \epsilon_z]$. Subsequently there is an error introduced in the RIR, namely,

$$h_\epsilon = \texttt{rir\_gen}(c, f_s, r_\epsilon, s, L, \beta, n) \tag{2}$$

Clearly an error in microphone position $\delta r = (r - r_\epsilon)$ results in an error in RIR namely $\delta h = h - h_\epsilon$. The problem we try to address is if there is a relationship between δr and δh.

Let us assume that the microphone position error δr is Gaussian with mean μ and variance σ^2, namely, $\mathcal{G}(\mu, \sigma^2)$. Note that typically sensor position error have a Gaussian distribution (see [29]). Does δh have any distribution? To understand the effect of the error made in measuring the receiver (microphone) location ($r = [x_r, y_r, z_r]$) on the RIR (h), we simulated the microphone position error ($\epsilon_x, \epsilon_y, \epsilon_z$) and computed the RIR as shown in Algorithm 1. Figure 1 shows the distribution of the microphone position error for $\sigma^2 = 0.1, 0.5,$ and 1. The distribution of the error in the RIR is shown in Figure 2 for both Euclidean distance (D_1) and Cosine distance (D_2). As can be seen the errors in both D_1 and D_2 show a distribution that is Gaussian ($\mathcal{G}(.,.)$). Note that the microphone position error has a mean ($\mu = 0$) however this results in a RIR error that has a non-zero mean.

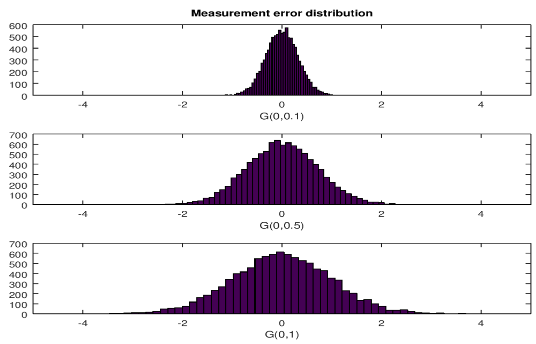

Figure 1: *Microphone position error ($\epsilon_x, \epsilon_y, \epsilon_z$) for $\sigma^2 = 0.1, 0.5, 1$.*

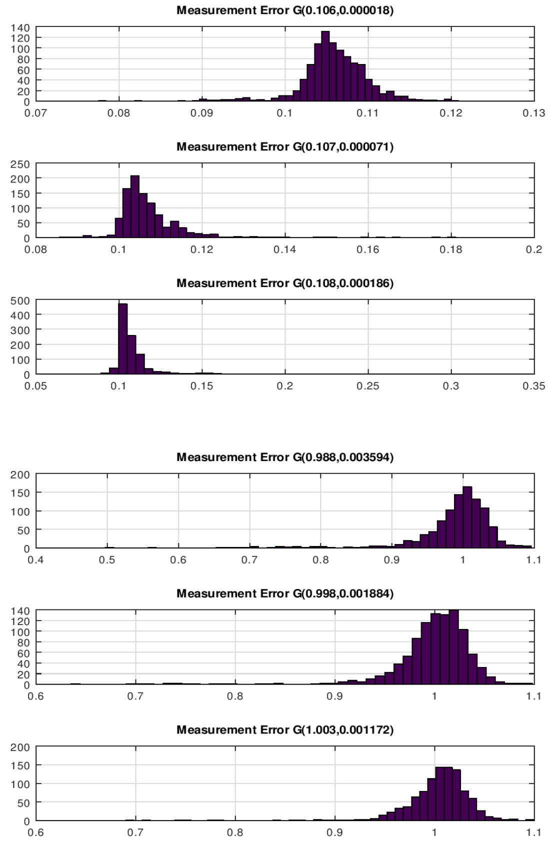

Figure 2: *RIR error D_1 (top) and D_2 (bottom) for measurement errors $\sigma^2 = 0.1, 0.5, 1$.*

Algorithm 1: Effect of microphone position error on RIR.

Input : $r = [x_r, y_r, z_r]$, c, f_s, s, L, β, n, $\texttt{rir\_gen}()$, σ^2, $\mathcal{G}(0, \sigma^2)$
Output: D_1 and D_2

for *iter = 1 to N* **do**
  Generate $\epsilon_x, \epsilon_y, \epsilon_z$ from $\mathcal{G}(0, \sigma^2)$
  Construct $r_\epsilon = [x_r + \epsilon_x, y_r + \epsilon_y, z_r + \epsilon_z]$.
  Compute $h = \texttt{rir\_gen}(c, f_s, r, s, L, \beta, n)$
  Compute $h_\epsilon = \texttt{rir\_gen}(c, f_s, r_\epsilon, s, L, \beta, n)$
  Find $D_1 = \text{EUCLIDEAN}(h, h_\epsilon)$
  Find $D_2 = \text{COSINE}(h, h_\epsilon)$
end
Procedure EUCLIDEAN(x, y)
return $\left(\sum_{i=1}^{n}(x_i - y_i)^2\right)$
procedure COSINE (x, y)
return $\left(\frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}\right)$

2.1. Signal Model

Room Impulse Response (RIR) can be estimated by transmitting a signal x from the loudspeaker and measuring the signal received at the microphone .There are some prior assumptions before defining the signal model. First, we assume that the initial K samples of RIR are the reasonable approximation of the whole RIR and second, that the signal x has unit variance. Let x be an $N \times 1$ training signal vector and y be the signal recorded at the microphone. The signal y can be modeled as (not considering the additive noise):

$$y(t) = h(t) \circledast x(t) \tag{3}$$

where $y(t)$ is the signal recorded at the microphone, $h(t)$ is the RIR and $x(t)$ is the transmitted signal. For RIR measurement, the closed microphone signal x is modeled as time shifted Toeplitz circular convolution matrix [30]. Signal recorded at microphone y now can be modeled as

$$y = Xh + a \tag{4}$$

where, X is a $N \times K$ matrix comprised of K shifted versions of x, h is the room impulse response vector of size $K \times 1$ and a is $K \times 1$ vector of additive ambient noise and is assumed to be independent to x.

2.2. Traditional RIR Measurement

Traditional way of estimating RIR [31] is by correlating the white training sequence x with the recorded speech y, as

$$\hat{h} = X^T y \tag{5}$$

After substituting the value of y from (4) into (5) we get

$$\hat{h} = X^T(Xh + a) = (X^T X)h + X^T a \tag{6}$$

Taking expectations on both sides of (6), we get

$$E\{\hat{h}\} = R_{xx}h + r_{xa} \tag{7}$$

where R_{xx} is the K-lag auto-correlation matrix ($E\{X^T X\}$) and r_{xa} is the K-lag cross-correlation vector ($E\{X^T a\}$). If signal sequence x is white then R_{xx} is I. Since noise a is assumed to be uncorrelated with x, r_{xa} will be 0. $X^T y$ is the unbiased estimate of h, increasing the length of signal $x(t)$, namely $N \rightarrow \infty$, improves the estimates of RIR (h). For finite signal sequence x, (7) can be approximated as:

$$\hat{h} = h + (X^T X - I)h + X^T a \tag{8}$$

As it can be observed that the predicted room impulse response (8) consists of two error components. The first error component is due to finite sequence signal correlation matrix not converging to the Identity matrix. The second component is due to the correlation vector between the sequence signal and the ambient noise, which are assumed to be independent. The error can be compensated in three ways, firstly by increasing the length of x, secondly increase the power (amplitude) of the signal, and thirdly by using the Maximum Length Sequence algorithm [32] [33] [34], which decreases the diagonal error due to auto-correlation matrix R_{xx}.

Table 1: *Full, Early and Late RIR variation on SDR with respect to change in microphone position*

	Full RIR			Early Part RIR			Late Part RIR		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
$1cm$	10.711	15.084	6.209	5.739	15.130	10.458	9.740	15.244	12.657
$2cm$	4.071	6.052	0.860	0.466	6.214	3.995	3.507	5.350	4.723
$3cm$	1.073	2.874	-1.242	-1.558	3.174	1.075	0.381	1.580	1.221
$4cm$	-0.022	1.928	-1.785	-2.025	2.284	-0.001	-0.821	0.390	0.005
$5cm$	-0.245	1.978	-2.059	-2.231	2.398	-0.022	-1.015	0.224	-0.191
$6cm$	-0.312	2.082	-2.223	-2.426	2.566	-0.296	-0.959	0.208	-0.204
$7cm$	-0.512	1.958	-2.289	-2.492	2.446	-0.505	-1.197	0.012	-0.345
$8cm$	-0.708	1.832	-2.307	-2.499	2.305	-0.715	-1.331	-0.075	-0.460
$9cm$	-0.772	1.852	-2.291	-2.482	2.301	-0.799	-1.097	0.015	-0.405
$10cm$	-0.765	1.923	-2.212	-2.412	2.381	-0.812	-0.969	0.064	-0.274

3. Impact of microphone position error on speech intelligibility and quality

We investigate the impact of RIR due to change in the microphone position on speech intelligibility and quality. In general, RIR comprises of early (including direct path) and late reflection components [35]. Their negative impacts on speech intelligibility have been studied in [36]. The early reflection predominantly creates the coloration effect resulting in each phonemes to have smearing of energies. On the other hand, the late reflections causes the degradation in low frequency envelopes of the speech signal due to overlap masking effect [36]. These early and late reflections have a significant detrimental effect on the perception of consonants and vowels which in turn lead to poor speech intelligibility and quality. We investigate the impact of change in full, early and late RIR on the intelligibility and quality of speech.

To best of our knowledge,

1. While studies have focused on understanding the impact of ambient noise on RIR, the effect of the microphone position error on RIR has not been studied.
2. The quantum of error in RIR (including both early and late) due to microphone position error has not been studied.
3. Methods to adress microphone position error in order to automatically tune the RIR for the current position also have not been looked into.

The first two points have been analysed in the earlier sections in this paper. This acts as a basis to tackle the third problem.

3.1. Microphone Position Error Analysis

The RIR estimation in an ambient noise is an open research problem and is a challenging task in a single microphone scenario. In order to identify the differences in the RIR measurement error due to ambient noise and due to change in position of microphone, we set an analogy between these two scenarios. As seen in (4), the signal received y can be modelled in the presence of ambient noise. Instead of ambient noise, if there is an error in measurement of microphone, then (4) can be modified as

$$\begin{aligned} y' &= Xh_\epsilon \\ &= X(h + \delta h) \\ &= Xh + X\delta h \end{aligned} \tag{9}$$

Where y' and h_ϵ are the recorded speech and erroneous RIR due to change in microphone position . The h is the RIR measured at the actual position of the microphone. As it can be observed from (9), there are two components, the actual signal captured at the microphone and the error term ($X\delta h$) introduced due to position error which is similar to the ambient noise a in (4). For the experimental analysis, we have simulated the room of dimensions 5m x 5m x 5m with source and microphone positions fixed at [3, 2 ,2] and [1, 2 , 2] respectively. The reverberation parameter RT_{60} is set to 0.3.

3.2. Impact of RIR variation on Speech Quality

The impact of position error is analysed in Table 1 in terms of variation in Signal-to-Distortion Ratio (SDR):

$$SDR = \sum_i \frac{(s[i])^2}{(s[i] - \hat{s}[i])^2} \tag{10}$$

where, s is the reference signal and $\hat{s}$ is the estimated signal. In Table 1, we consider the error distance between 1 cm to 10 cm. The error analysis shows that even with a small microphone position shift of 1 cm causes the SDR of 10.711 dB, which in analogy is equivalent to SNR of 10 dB (ambient noise). This kind of SNR in speech signal processing, specifically the performance

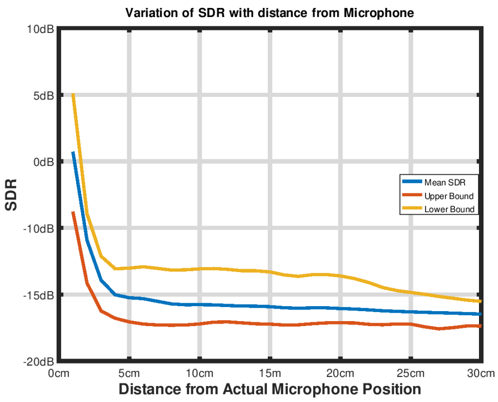

Figure 3: *SDR variation with different microphone positions.*

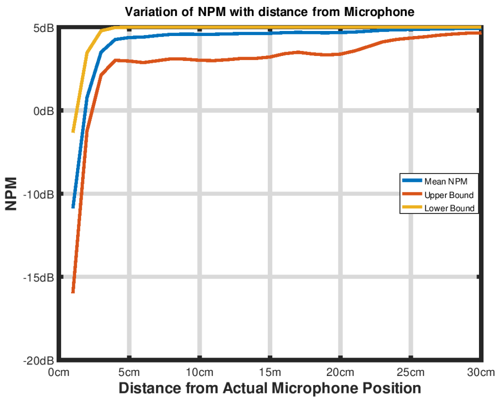

Figure 4: *NPM variation with different microphone positions.*

of automatic speech recognition degrades significantly for signals of 10 dB SNR. On an average, there is a loss of 5 dB, 6 dB and 8 dB in SDR for full, early and late RIR respectively, when microphone position is changed from 1 cm to 2 cm. Any increase in the microphone position, the loss in SDR increases gradually. Figure 3 gives the bounds of SDR or sensitivity of RIR upto 30 cm position displacement error. The key observation here is an abrupt decrease in the SDR upto 5 cm and then it remains constant for increase in measurement error. A displacement of, as little as 5 cm causes SDR to drop to -0.245 dB.

Similarly, the normalized projection misalignment (NPM) scores were computed between the actual RIR and the estimated RIR at various error positions (see Figure 4). A similar trend of abrupt change upto 5 cm can be observed, indicating that even small errors in the microphone positions can introduce very high distortions. This in turn indicates that the quality of speech undergoes serious distortions with slight change in microphone position because of the change in RIR. However, the impact of early and late RIR variation on NPM is more or less similar to full RIR, hence not presented herein.

3.3. Impact of RIR on Speech Intelligibility

Speech Intelligibility maps to the amount of information that can be perceived by a listener. It is measured using short time objective intelligibility (STOI) [37] [38], which gives a score between 0 and 1 (higher the better). Figure 5 shows the variation of STOI scores with different microphone positions, where the actual position is taken as reference. We can observe in Figure 5 that even a small shift in the position of the microphone there is a significant change in STOI and hence degradation in speech intelligibility is also significant. Similar observations related to early and late RIR effect on speech intelligibility are visible in Figure 5. However, the STOI for late RIR is most detrimental in comparison to other two. But STOI variation for late RIR, with change in microphone positions are not significant as compared to early and full RIR's.

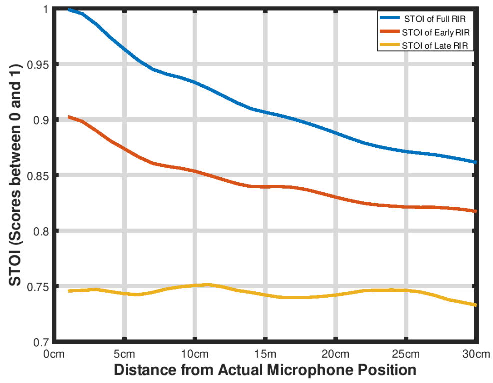

Figure 5: *STOI variation with varying microphone positions.*

4. Conclusions

In this paper, we showed that an error in measuring the position of the microphone can have a significant influence on the RIR. We further showed that if the measurement error has a Gaussian distribution, the error in RIR also showed a similar distribution with a shift in the mean. To bring the study into perspective, we showed that the RIR error due to ambient noise, which has been studied in the literature, is comparable to the error introduced by microphone position error. Our studies show that even in the absence of ambient noise, a position error in microphone, of 1 cm, introduces an error equivalent to ambient noise of 10 dB. This correspondence, in our opinion, has not been studied before. We showed that there is a drastic error found in SDR, Normalized Projection Misalignment (NPM) and short time objective intelligibility (STOI) scores, when there is a change in microphone position. This shows the impact of microphone position measurement error on both quality and intelligibility of speech.

With speech solutions taking center stage, speech recognition in the presence of degraded speech is getting prominence, we believe this will lead to dynamic computing of RIR especially in the presence of measurement error and ambient noise. A considerable amount of attention to computing RIR in the presence of microphone position measurement errors is to be expected in a near future with commercialization of smart speakers like Google Home, Alexa etc.

5. References

[1] P. A. Naylor and N. D. Gaubitch, *Speech dereverberation*. Springer Science & Business Media, 2010.

[2] 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.

[3] W. C. Sabine, *Collected papers on acoustics*. Harvard University Press Cambridge, MA, 1927.

[4] C. F. Eyring, "Reverberation time in dead rooms," *The Journal of the Acoustical Society of America*, vol. 1, no. 2A, pp. 217–241, 1930.

[5] E. A. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 770–773, 2009.

[6] T. Betlehem and T. D. Abhayapala, "A modal approach to soundfield reproduction in reverberant rooms," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 3. IEEE, 2005, pp. iii–289.

[7] M. Crocco, A. Trucco, V. Murino, and A. Del Bue, "Towards fully uncalibrated room reconstruction with sound," in *2014 22nd European Signal Processing Conference (EUSIPCO)*. IEEE, 2014, pp. 910–914.

[8] A. H. Moore, M. Brookes, and P. A. Naylor, "Room geometry estimation from a single channel acoustic impulse response," in *21st European Signal Processing Conference (EUSIPCO 2013)*. IEEE, 2013, pp. 1–5.

[9] T. Yoshioka, A. Sehr, M. Delcroix, K. Kinoshita, R. Maas, T. Nakatani, and W. Kellermann, "Making machines understand us in reverberant rooms: Robustness against reverberation for automatic speech recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 114–126, 2012.

[10] A. Krueger and R. Haeb-Umbach, "Model-based feature enhancement for reverberant speech recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1692–1707, 2010.

[11] L. L. Beranek and T. Mellow, *Acoustics: sound fields and transducers*. Academic Press, 2012.

[12] A. Muthukumarasamy, "Impact of microphone positional errors on speech intelligibility," Master's thesis, University of Kentucky Master's Theses, Lexington, Kentucky, 2009, masters Thesis 602.

[13] J. M. Sachar, H. F. Silverman, and W. R. Patterson, "Position calibration of large-aperture microphone arrays," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2. IEEE, 2002, pp. II–1797.

[14] I. Szöke, M. Skácel, L. Mošner, J. Paliesek, and J. H. Černockỳ, "Building and evaluation of a real room impulse response dataset," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 4, pp. 863–876, 2019.

[15] M. Crocco and A. Del Bue, "Room impulse response estimation by iterative weighted l 1-norm," in *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, pp. 1895–1899.

[16] B. Dumortier and E. Vincent, "Blind rt60 estimation robust across room sizes and source distances," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 5187–5191.

[17] Y. Lin, J. Chen, Y. Kim, and D. D. Lee, "Blind sparse-nonnegative (bsn) channel identification for acoustic time-difference-of-arrival estimation," in *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE, 2007, pp. 106–109.

[18] K. Kowalczyk, E. A. Habets, W. Kellermann, and P. A. Naylor, "Blind system identification using sparse learning for tdoa estimation of room reflections," *IEEE Signal Processing Letters*, vol. 20, no. 7, pp. 653–656, 2013.

[19] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Transactions on signal processing*, vol. 51, no. 1, pp. 11–24, 2003.

[20] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE assp magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[21] J. M. Sachar, H. F. Silverman, and W. R. Patterson, "Microphone position and gain calibration for a large-aperture microphone array," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 42–52, 2004.

[22] D. Florencio and Z. Zhang, "Maximum a posteriori estimation of room impulse responses," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 728–732.

[23] A. Raikar, S. Basu, and R. M. Hegde, "Single channel joint speech dereverberation and denoising using deep priors," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2018, pp. 216–220.

[24] S. Mosayyebpour, A. Sayyadiyan, M. Zareian, and A. Shahbazi, "Single channel inverse filtering of room impulse response by maximizing skewness of LP residual," in *2010 International Conference on Signal Acquisition and Processing*. IEEE, 2010, pp. 130–134.

[25] H. Padaki, K. Nathwani, and R. M. Hegde, "Single channel speech dereverberation using the LP residual cepstrum," in *2013 National Conference on Communications (NCC)*. IEEE, 2013, pp. 1–5.

[26] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[27] E. A. Habets, "Room impulse response generator," *Technische Universiteit Eindhoven, Tech. Rep*, vol. 2, no. 2.4, p. 1, 2006.

[28] ——, *RIR-Generator*, 2016. [Online]. Available: https://github.com/ehabets/RIR-Generator

[29] S. Kopparapu and P. Corke, "The effect of noise on camera calibration parameters," *Graph. Models*, vol. 63, no. 5, p. 277303, Sep. 2001. [Online]. Available: https://doi.org/10.1006/gmod.2001.0551

[30] G. Strang, "A proposal for toeplitz matrix calculations," *Studies in Applied Mathematics*, vol. 74, no. 2, pp. 171–176, 1986.

[31] G. Turin, "An introduction to matched filters," *IRE transactions on Information theory*, vol. 6, no. 3, pp. 311–329, 1960.

[32] J. Vanderkooy, "Aspects of mls measuring systems," *Journal of the Audio Engineering Society*, vol. 42, no. 4, pp. 219–231, 1994.

[33] I. Mateljan, "Signal selection for the room acoustics measurement," in *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99 (Cat. No. 99TH8452)*. IEEE, 1999, pp. 199–202.

[34] E. A. Habets, "Multi-channel speech dereverberation based on a statistical model of late reverberation," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 4. IEEE, 2005, pp. iv–173.

[35] J. S. Bradley, H. Sato, and M. Picard, "On the importance of early reflections for speech in rooms," *The Journal of the Acoustical Society of America*, vol. 113, no. 6, pp. 3233–3244, 2003.

[36] Y. Hu and K. Kokkinakis, "Effects of early and late reflections on intelligibility of reverberated speech by cochlear implant listeners," *The Journal of the Acoustical Society of America*, vol. 135, no. 1, pp. 22–28, 2014.

[37] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *2010 IEEE international conference on acoustics, speech and signal processing*. IEEE, 2010, pp. 4214–4217.

[38] ——, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, 2011.