Performance Analysis for Blind Identification of Acoustic Channels

Majid Fozunbal

HP Laboratories
HPL-2009-333

Keyword(s):
Audio, acoustic systems, blind identification, dereverberation, deconvolution, multichannel

Abstract:
Multichannel blind system identification is a prominent part in audio dereverberation. Despite much progress in this field, the performance of existing algorithms is still unsatisfactory and existing theories have failed to properly explain the issue. To assess performance impediments, we introduce, quantify, and illustrate three major sources of errors: namely, estimation, approximation, and referencing errors. For each error, we provide simple expressions that describe the effect of various system parameters and that can serve as a guideline to explain reality and improve performance.

External Posting Date: October 6, 2009 [Fulltext] Approved for External Publication
Internal Posting Date: October 6, 2009 [Fulltext]

PERFORMANCE ANALYSIS FOR BLIND IDENTIFICATION OF ACOUSTIC CHANNELS

Majid Fozunbal

Hewlett-Packard Laboratories
1501 Page Mill Road
Palo Alto, CA 94304

ABSTRACT

Multichannel blind system identification is a prominent part in audio dereverberation. Despite much progress in this field, the performance of existing algorithms is still unsatisfactory and existing theories have failed to properly explain the issue. To assess performance impediments, we introduce, quantify, and illustrate three major sources of errors: namely, *estimation*, *approximation*, and *referencing* errors. For each error, we provide simple expressions that describe the effect of various system parameters and that can serve as a guideline to explain reality and improve performance.

Index Terms— Acoustic systems, blind identification, dereverberation, deconvolution, multichannel.

1. INTRODUCTION

Multichannel blind identification or estimation is a central problem in many audio processing algorithms including dereverberation and source separation [1]. A common approach to dereverberation is a two-stage method composing a blind identification stage followed by an equalization stage implemented by inverse filtering [1]. Unlike non-blind techniques that are used in echo control applications, blind identification techniques do not have a clean and separate reference for adaptation purposes. Instead, they employ adaptive strategies in which each channel benchmarks its performance against those of other channels [1]. Thus, blind techniques are not only susceptible to the same impairments of non-blind methods, but also prone to cross channel referencing errors [2].

Blind identification and equalization techniques have a rich history ingrained in communication systems. Yet, their applications in acoustic systems have been hindered by the large dimensionality of acoustic channels. As a result, existing blind techniques have not been successful to serve as a robust base for audio dereverberation [3]. Recently, there have has been some progress to address the issue. In [4], the authors propose a heuristic approach to enhance robustness against additive noise. They, however, did not address the intrinsic causes of susceptibility. Consequently, in [3], the authors attribute performance sensitivity to near-common zero of room impulses and propose an approach to improve performance [5]. Despite this progress, a comprehensive study to dissect sources of error, to assess their levels, and to provide insights on how to tackle them is missing. In this spirit, this paper attempts to analyze the performance of multichannel blind estimation in a holistic, analytical approach. More precisely, we introduce three major sources of error in optimization of blind estimation: *estimation*, *approximation*, and *referencing* errors. Estimation error is caused by empirical data. Using arguments from statistical learning theory, we assess and provide a simple upper bound. In contrast, approximation error is characterized as a mismatch between the length of room impulse responses, L, and estimation filters, $l \leq L$. Deriving a simple expression, we quantify and numerically illustrate approximation error. Moreover, we illustrate the trade-off between these two errors with respect to l.

Finally, we address referencing error that is caused by cross referencing. In theory, if there exists no common-zero among channel impulses, then blind identification algorithms converge to the true solution [1]. In practice, however, this condition does not suffice convergence. To investigate the reason, we introduce a measure of sensitivity in optimization. Augmenting existing theories [3], our results provide new insights in predicting performance.

2. SETUP

Consider a conference room with two microphones and an audio source. Let $s(n)$ denote the audio signal that is generated by the source. The acoustic coupling between the source and microphone i is called *reverb path* and is denoted by a filter (vector) $h_i \in \mathbb{R}^L$, in which L is the *reverb length*. Fig. 1 depicts two such reverb paths for a sampling frequency of 8KHz. Using a linear model, microphone signals are described by

$$x_i(n) = (h_i * s)(n) + w_i(n), \ i = 1, 2 \tag{1}$$

where $w_i(n)$ denotes the background noise recorded by microphone i. The noise is considered temporally and spatially uncorrelated and independent from $s(n)$. Observing $x_i(n)$, the goal of dereverberation is to find two filters f_i such that

$$y(n) = \sum_{i=1}^{2} (f_i * x_i)(n) \propto s(n). \tag{2}$$

A common approach to this problem is based on blind estimation of h_i followed by inverse filtering using the obtained estimates. Hence, any inaccuracy in estimation manifests it-

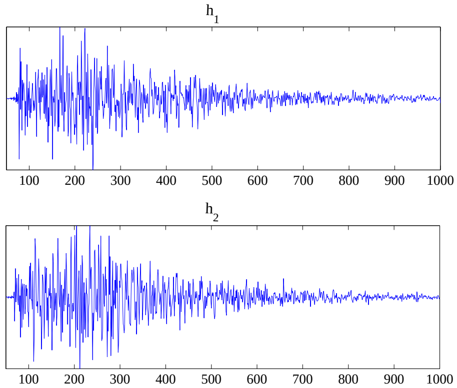


Fig. 1. Typical reverb paths recorded in a conference room for a sampling frequency of 8KHz. The microphones are 3 feet apart and the source is 5 feet from their center.

self into a large deviation in inversion.

Blind identification techniques estimate the reverb paths through cross referencing by minimizes the *empirical risk*

$$\hat{J}_m(\boldsymbol{g}) = \hat{\mathbb{E}}_m \left\{ \|g_1 * x_1(n) - g_2 * x_2(n)\|^2 \right\} \quad (3)$$

for *estimation filters* $\boldsymbol{g} = (g_1, g_2) \in \mathcal{G}$ where

$$\mathcal{G} = \{\boldsymbol{g} \in \mathbb{R}^{2l} : \|\boldsymbol{g}\| = 1, \mathbf{1}'\boldsymbol{g} \geq 0\} \quad (4)$$

is the *estimation space*. Here, $l \leq L$ is the *estimation length*, i.e., the length of g_1 and g_2. The subscript m denotes the sample size, i.e., the size of empirical data, used in computing empirical expectation (3). Moreover, the condition $\mathbf{1}'\boldsymbol{g} \geq 0$ in Eq. (4) is to exclude trivial sign symmetric solutions.

Theoretically, it is known that if the z-transform of h_1 and h_2 have no common-zeros and if the estimation length, l, is exactly L, then as $m \to \infty$, g_1 and g_2 converge–uniquely up to a scalar–to h_2 and h_1, respectively. This is the sufficient condition that is often cited as *no common-zero condition* [3]. In practice, despite the fact that this condition holds for room impulse responses, like those shown in Fig. 1, the optimization (3) of blind estimators fails to accurately estimate h_2 and h_1. The inherent frailty and vulnerability in minimizing Eq. (3) is a fundamental challenge that has left blind estimation for dereverberation still open.

3. ERROR ANALYSIS

There are three major sources of error in optimization (3) that hinder the performance of blind estimation algorithms: *estimation error*, *approximation error*, and *referencing error*.

3.1. Estimation Error

The estimation error is simply defined as the error caused by using the empirical risk (3) instead of the true risk

$$J(\boldsymbol{g}) = \mathbb{E} \left\{ \|g_1 * x_1(n) - g_2 * x_2(n)\|^2 \right\}. \quad (5)$$

Intuitively, since $x_i(n)$ is random, at each point $\boldsymbol{g}$, $\hat{J}_m(\boldsymbol{g})$ is a random variable that deviates from its mean $J(\boldsymbol{g})$. This deviation is the source of estimation error.

To assess the level of estimation error, assume that $|x_i(n)| < B$. Using the principle of empirical risk minimization (ERM) [6, Theorem 5.1], we have

$$J(\boldsymbol{g}) \leq \hat{J}_m(\boldsymbol{g}) + 2B^2\sqrt{\varepsilon(m,l)} \quad (6)$$

for sufficiently large values of m and with a probability no smaller than the *confidence factor* $1-\beta$. The second term in Eq. (6) is

$$\varepsilon(m,l) = \frac{1}{m}\left[2l(\ln\frac{m}{l}+1) - \ln\beta/4 + 1\right] \quad (7)$$

that determines an upper-bound on pointwise deviation of the risks, a deviation that converges to zero in probability as $m \to \infty$.

Suppose $\boldsymbol{g}^m = (g_1^m, g_2^m)$ and $\boldsymbol{g}^o = (g_1^o, g_2^o)$ minimize the empirical risk (3) and the true risk (5), respectively. The true risk penalty incurred by using $\boldsymbol{g}^m$ instead of $\boldsymbol{g}^o$ is bounded by

$$J(\boldsymbol{g}^m) - J(\boldsymbol{g}^o) \leq 2B^2\left(\sqrt{\varepsilon(m,l)} + \sqrt{-\frac{\ln\beta}{2m}}\right) \quad (8)$$

with probability no smaller than $1-2\beta$ [6, pp. 193].

Assume the source and noise signals are uncorrelated, zero mean, and white with variances σ_s^2 and σ_w^2, respectively. We can show that

$$J(\boldsymbol{g}) = \sigma_s^2 D_{\boldsymbol{h}}(\boldsymbol{g}) + \sigma_w^2$$

in which

$$D_{\boldsymbol{h}}(\boldsymbol{g}) = \|g_1 * h_1 - g_2 * h_2\|^2 \quad (9)$$

is the *misalignment* in estimating $\boldsymbol{g} = (g_1, g_2)$ for the underlying $\boldsymbol{h} = (h_1, h_2)$. We define the *estimation error* as the additional misalignment incurred by using $\boldsymbol{g}^m$ instead of $\boldsymbol{g}^o$. Rewriting Eq. (8) in terms of $D_{\boldsymbol{h}}(\boldsymbol{g})$, we obtain the bound

$$D_{\boldsymbol{h}}(\boldsymbol{g}^m) - D_{\boldsymbol{h}}(\boldsymbol{g}^o) \leq \overline{\text{Er}}(\text{est})$$

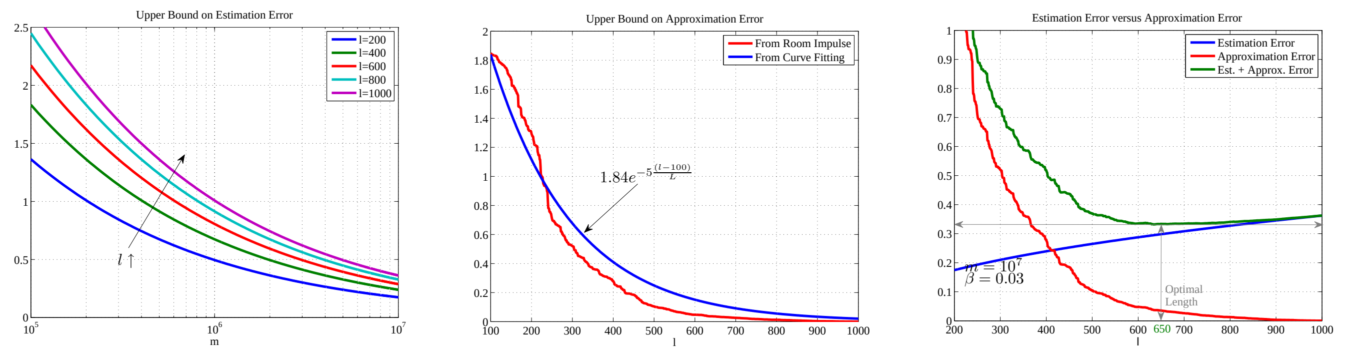
$$\overline{\text{Er}}(\text{est}) \triangleq 8(1+\frac{1}{\text{SNR}})\left(\sqrt{\varepsilon(m,l)} + \sqrt{-\frac{\ln\beta}{2m}}\right) \quad (10)$$

that holds true with a probability of at least $1-2\beta$. The bound (10) is a function of sample size, m, estimation length, l, and signal to noise ratio, $\text{SNR} = \frac{\sigma_s^2}{\sigma_w^2}$. In derivation of (10), we adjusted Eqs. (6) and (8) for Gaussian signals by taking B as two times the standard deviation with a confidence of 97%. For $\|\boldsymbol{h}\| = 1$, this means that $B^2 = 4(\sigma_s^2 + \sigma_w^2)$.

Fig. 2(a) depicts the behavior of the bound (10) for five different estimation lengths and for $\beta = 0.03$. The bound illustrates that, for fixed l, as m increases, the estimation error diminishes. In contrast, for fixed m, as l increases, the error increases.

3.2. Approximation Error

The estimation error is caused by minimizing empirical risk instead of the true risk. In contrast, approximation error is due to limiting the search space $\mathcal{G}$ when minimizing $D_{\boldsymbol{h}}(\boldsymbol{g})$.



(a) (b) (c)

Fig. 2. (a) Estimation error for five estimation lengths with $\beta = 0.03$. (b) Approximation error for a collection of room impulses, for which $L = 1000$, measured in the same room as in Fig. 1. (c) Estimation + Approximation error suggesting $l = 650$ as an optimal estimation length.

More precisely, the *approximation error* is defined by

$$\mathrm{Er(app)} = \sup_{\boldsymbol{h}\in\mathcal{H}} \min_{\boldsymbol{g}\in\mathcal{G}} D_{\boldsymbol{h}}(\boldsymbol{g}). \tag{11}$$

Finding the exact expression of the approximation error is difficult or even impossible depending on definition of $\mathcal{H} \subset \mathbb{R}^{2L}$. For blind estimation of room impulses, however, we find a simple upperbound on approximation error (11) as follows.

From Fig. 1, we note that the energy of reverberation paths drops as a function of time. Thus, we model $\mathcal{H}$ as a collection of $\boldsymbol{h} = (h_1, h_2)$ such that $\sum_i \|h_i\|^2 = 1$ and both h_1 and h_2 have diminishing amount of energy in their tails. For each estimation length $l \leq L$ and for each $h = (a_1, \cdots, a_L)$, let

$$h^{(l)} \triangleq (a_1, \cdots, a_l, \underbrace{0, \cdots, 0}_{L-l})$$

denotes the truncated version of h. Moreover, let

$$\gamma(l) = \sup_h \|h - h^{(l)}\|^2 \tag{12}$$

denote the maximum energy of the tail $L - l$ elements among all h. With some routine application of triangular inequality, we derive

$$\overline{\mathrm{Er}}(\mathrm{app}) \triangleq 2\gamma(l) \tag{13}$$

as an upperbound on approximation error. Fig. 2(b) depicts this bound for a collection of impulse responses that were obtained from the conference room whose sample impulses are shown in Fig. 1. Fig. 2(b) also depicts an exponential curve that is fit to $\gamma(l)$ suggesting a parametric upper bound

$$\overline{\mathrm{Er}}(\mathrm{app}) = \rho e^{-\alpha l}, \;\; \rho > 0, \alpha > 0, \tag{14}$$

that simplifies further analytical derivations.

To better understand the dynamics of estimation error (10) versus approximation error (13), Fig. 2(c) depicts them together to contrast their opposite behavior with respect to estimation length, l. As shown, the sum of these errors indicates that $l = 650$ is the optimal estimation length; its analytical expression can be obtained by minimizing the sum of (10) and (13) with respect to l.

3.3. Referencing Error

We note that neither of the aforementioned bounds depend on the specific underlying $\boldsymbol{h}$. Instead, they depend on its generic characteristics such as unit norm and tail behavior. In contrast, there is another source of error that is specific to underlying $\boldsymbol{h}$. We call this referencing error because in minimizing the misalignment (9), choices for g_1 and g_2 are prone to one another's errors. To assess the inherent vulnerability for such error, we use the notion of radius of uncertainty [7]. That is for every $\epsilon > 0$, let

$$\mathcal{G}_{\boldsymbol{h}}(\epsilon) = \{\boldsymbol{g} \in \mathcal{G} : D_{\boldsymbol{h}}(\boldsymbol{g}) \leq \min_{\boldsymbol{g}\in\mathcal{G}} D_{\boldsymbol{h}}(\boldsymbol{g}) + \epsilon^2\} \tag{15}$$

as the ϵ-solution set for $\boldsymbol{h}$. The radius of uncertainty is defined

$$r_{\boldsymbol{h}}(\epsilon) = \sup_{g_1, g_2 \in \mathcal{G}_{\boldsymbol{h}}(\epsilon)} \frac{\|\boldsymbol{g}_1 - \boldsymbol{g}_2\|}{2}. \tag{16}$$

We may assume that $L = l$. Hence, for every $\boldsymbol{h}$, there exists an optimal solution $\boldsymbol{g}^o = (g_1^o, g_2^o) \in \mathcal{G}_{\boldsymbol{h}}(\epsilon)$ that exactly estimates $\boldsymbol{h}$ up to a scalar. By Taylor expansion at $\boldsymbol{g}^o$, we have

$$\mathcal{G}_{\boldsymbol{h}}(\epsilon) = \{\boldsymbol{g} \in \mathcal{G} : \boldsymbol{g} = \boldsymbol{g}^o + \boldsymbol{v}, \;\; \boldsymbol{v}'\nabla^2 D_{\boldsymbol{h}}(\boldsymbol{g}^o)\boldsymbol{v} \leq \epsilon^2\}. \tag{17}$$

Assuming the no common-zero condition on h_1 and h_2, $\mathcal{G}_{\boldsymbol{h}}(\epsilon)$ lies on an ellipsoid that has no direction of degeneration, excluding $\boldsymbol{g}^o$. Hence, the radius of uncertainty is bounded by the largest axis of this ellipsoid. Thus, we have

$$\frac{r_{\boldsymbol{h}}(\epsilon)}{\epsilon} \leq \frac{1}{\sqrt{\sigma_2(\nabla^2 D_{\boldsymbol{h}}(\boldsymbol{g}^o))}}. \tag{18}$$

as a measure of vulnerability in estimating $\boldsymbol{h}$, aligned with heuristic results of [3]. In Eq. (18), $\sigma_2(\cdot)$ denotes the smallest non-zero eigenvalue of the Hessian matrix

$$\nabla^2 D_{\boldsymbol{h}}(\boldsymbol{g}^o) = \begin{bmatrix} H_1' \\ -H_2' \end{bmatrix} \begin{bmatrix} H_1 & -H_2 \end{bmatrix} \tag{19}$$

where H_1 and H_2 denote $(2l-1) * l$ Toeplitz matrices derived from h_1 and h_2, respectively.

The Hessian matrix is of size $2l$. By no common-zero condition, it has the maximum rank of $2l - 1$ and has a one dimensional null space corresponding to the direction of $\boldsymbol{g}^o$. The goal of blind estimation is to find this eigenvector. But, its performance is hindered by other small eigenvectors of Hessian matrix. Hence, such blind techniques are extremely sensitive to any computation error. This sensitivity, whose severity can be measured by (18), manifests itself as either instability in estimation or an extremely slow convergence making the algorithms practically useless.

Fig. 3 depicts the eigenvalues of this matrix for two types of impulse responses: 1) room impulses responses shown in Fig. 1, and 2) randomly generated impulse responses. In each case, the length of the impulse responses is 1000 and the rank of the matrix is 1999, the maximum possible rank. However, it is seen that a majority of eigenvalues are extremely small. Although, impulse responses have no common-zeros, blind estimation algorithms fail to estimate $\boldsymbol{h}$. This is more severe for room impulse responses, as it can be inferred from Fig. 3 and it can be predicated by Eq. (18).

We do not have a simple description for Eq. (18). However, to have a practical rule of thumb, we may find an approximation for Eq. (18) by approximating Toeplitz matrices, H_i, with circulant matrices. As a result, we obtain the approximate similarity

$$\nabla^2 D_{\boldsymbol{h}}(\boldsymbol{g}^o) \sim \begin{bmatrix} D|\mathcal{F}h_1|^2 & -D\mathcal{F}h_1^* \circ \mathcal{F}h_2 \\ -D\mathcal{F}h_2^* \circ \mathcal{F}h_1 & D|\mathcal{F}h_2|^2 \end{bmatrix} \quad (20)$$

where $\mathcal{F}h$ denotes the discrete fourier transform of h, '$\circ$' denotes the pairwise multiplication, and D maps a vector to a diagonal matrix. Half of the eigenvalues of this matrix are 0 and the other half are

$$|\mathcal{F}h_1(k)|^2 + |\mathcal{F}h_2(k)|^2, \quad k \in \{1, \cdots, l\} \quad (21)$$

where k denotes the frequency bins. Hence, the smallest non-zero eigenvalue of the approximate Hessian Matrix is

$$\min_k \left(|\mathcal{F}h_1(k)|^2 + |\mathcal{F}h_2(k)|^2\right), \quad (22)$$

which is the median of all of its eigenvalues. Thus, we may use

$$\frac{r_{\boldsymbol{h}}(\epsilon)}{\epsilon} \approx \Omega\left(\frac{1}{\min_k \sqrt{|\mathcal{F}h_1(k)|^2 + |\mathcal{F}h_2(k)|^2}}\right) \quad (23)$$

as a rule of thumb measure of sensitivity in optimization (3). Eq. (23) is an augmentation to the no common-zero sufficient condition. It sheds light on sensitivity and frailty of the performance of blind estimation.

4. CONCLUSION

This paper attempts to provide a holistic, analytical interpretation and assessment of practical impairments of multichannel blind estimation techniques. We introduced and quantified three major sources of error and illustrated their effects on performance. The results of this work just scratch the tip of the iceberg in fully understanding the sensitive dynamics of blind estimation. Yet, they can serve as a guideline for further detailed analysis and improved design techniques.

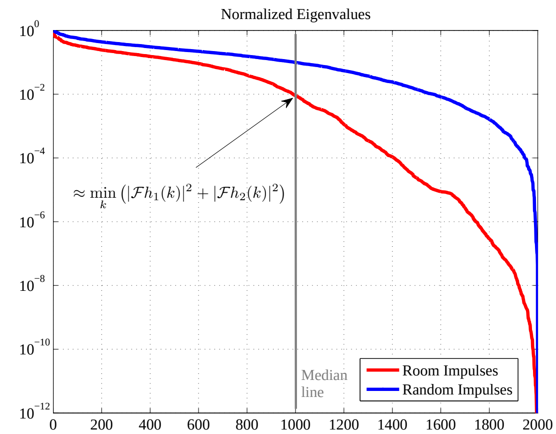


Fig. 3. Normalized Eigenvalues of Hessian matrix (19) for two types of impulse responses: room impulses in Fig. 1 and random impulses. In each case, $L = 1000$ and the rank of Hessian matrix is 1999 implying no common-zero condition.

ACKNOWLEDGMENT

The author is grateful to Mitch Trott for stimulating discussions that kindled this work. He is also thankful to Ton Kalker for editorial comments that improved the presentation of the paper.

5. REFERENCES

[1] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Trans. Speech Audio Process.*, vol. 13, pp. 882–895, 2005.

[2] H.H. Yanga and Y. Huab, "On performance of cross-relation method for blind-channel identification," *Elsevier Signal Processing*, vol. 62, pp. 187–205, 1997.

[3] P.A. Naylor, X. Lin, and A.W.H. Khong, "Near-common zeros in blind identification of SIMO acoustic systems," *Proc. IEEE HSCMA*, 2008.

[4] N. D. Gaubitch, M. K. Hasan, and P. A. Naylor, "Noise robust adaptive blind channel identification using spectral constraints," *Proc. IEEE ICASSP*, May 2006.

[5] A.W.H. Khong, X. Lin, and P.A. Naylor, "Algorithms for identifying clusters of near-common zeros in multichannel blind system identification and equalization," *Proc. IEEE ICASSP*, Apr. 2008.

[6] Vladimir Vapnik, *Statistical Learning Theory*, John Wiley & Sons, 1998.

[7] J. F. Traub, G. W. Wasilowski, and H. Woznikowski, *Information, Uncertainty, Complexity*, Addison Wesley, 1983.