

EXPECTED PERFORMANCE OF A FAMILY OF BLIND SOURCE SEPARATION ALGORITHMS IN A REVERBERANT ROOM

Fotios Talantzis, Darren B. Ward and Patrick A. Naylor

Dept. of Electrical and Electronic Engineering, Imperial College London, SW7 2BT, UK

ABSTRACT

Using statistical room acoustics we investigate the performance of Blind Source Separation and Deconvolution (BSSD) algorithms when used in a reverberant room. We focus on the case where one of the sources moves, and examine the relative impact of source movement and room reverberation on the expected performance. We derive theoretical expressions, and verify these through image model simulations.

1. INTRODUCTION

Blind Source Separation and Deconvolution (BSSD) is a technique that has gained extensive attention recently, since it presents us with a potentially powerful set of tools to separate and deconvolve speech signals when only mixed observations are available.

Unfortunately, performance in real acoustic environments is limited by the unpredictability of room reverberation. In particular, if the sources move, the performance of the system may degrade significantly. This problem has been investigated to some extent using BSSD experiments of speech phoneme recognition in [2], where it was found that source movement degraded the recognizer performance. It has been noted that if the BSSD system is constrained to only separate the direct-path components of the mixture, fast algorithm convergence is obtained [4], but performance suffers as the enclosure becomes more reverberant.

The performance of BSSD algorithms has typically been measured using Signal-to-Noise Ratios (SNR) at the input and output of the system. This requires knowledge of the source signals and mixtures (i.e. the characteristics of the enclosure). These measures are dependent on the particular test environment used for each BSSD algorithm, making generalized conclusions difficult to formulate.

Our aim in this paper is to investigate in more detail the effect of source movement and reverberation on BSSD algorithms, by leveraging results from statistical room acoustics (SRA). SRA assumes that the energy density of the sound waves inside an enclosure can be considered as the sum of contributions from an infinite number of plane waves that arrive from random directions of propagation and contribute equally to the space-averaged energy density. In [3] we used

SRA to examine the performance of multi-channel acoustic equalization systems. Here, we extend our results to the specific case of BSSD with two sources and two microphones.

2. PROBLEM SPECIFICATION

2.1. Signal Model

Consider the case of two source signals impinging on two microphones. Let the position of the first and second sources, respectively, be $\mathbf{s}_1$ and $\mathbf{s}_2$. Similarly, let the position of the first and second microphone be $\mathbf{m}_1 = [x_1, y_1, z_1]^T$ and $\mathbf{m}_2 = [x_2, y_2, z_2]^T$, respectively. Since SRA is naturally formulated in the frequency domain, we will consider a frequency-domain formulation of the source separation problem.

Define $A(\mathbf{s}, \mathbf{m})$ as the acoustic transfer function (ATF) from a source located at $\mathbf{s}$ to a microphone located at $\mathbf{m}$. This ATF can be written as the sum of direct-path $A_d(\mathbf{s}, \mathbf{m})$ and reverberant-path $A_r(\mathbf{s}, \mathbf{m})$ terms as

$$A(\mathbf{s}, \mathbf{m}) = A_d(\mathbf{s}, \mathbf{m}) + A_r(\mathbf{s}, \mathbf{m}), \quad (1)$$

where the direct-path term is

$$A_d(\mathbf{s}, \mathbf{m}) \triangleq \frac{e^{jk\|\mathbf{s}-\mathbf{m}\|}}{4\pi\|\mathbf{s}-\mathbf{m}\|}, \quad (2)$$

where $k = 2\pi f/c$ is the wavenumber with c the speed of sound propagation in air, f the frequency and $\|\cdot\|$ is the vector 2-norm.

Let x_1 and x_2 denote the signals at the first and second microphones, respectively. These are given by

$$\mathbf{x} = \mathbf{A}\mathbf{u}, \quad (3)$$

where $\mathbf{x} = [x_1, x_2]^T$ is the vector of microphone signals, $\mathbf{u} = [u_1, u_2]^T$ is the vector of source signals, and $\mathbf{A}$ is the 2×2 matrix of ATFs, with the (n, k) th element given by $[\mathbf{A}]_{n,k} = A(\mathbf{s}_k, \mathbf{m}_n)$. Note that all signals and transfer functions are functions of frequency, although to simplify notation we suppress the explicit dependence on frequency.

These microphone signals are then filtered by the 2×2 unmixing matrix $\mathbf{H}$ to give the output signals:

$$\mathbf{y} = \mathbf{H}\mathbf{x}, \quad (4)$$

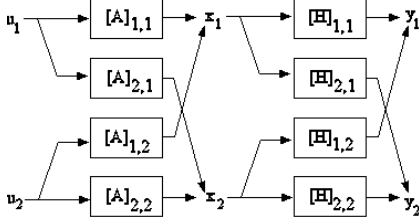


Fig. 1. Block diagram of the standard BSS system.

where $\mathbf{y} = [y_1, y_2]^T$ is the vector of output signals. Let $H_n(\mathbf{m}_k)$ denote the unmixing filter from the microphone at $\mathbf{m}_k$ to the n th output. Hence, the (n, k) th element of the unmixing matrix is $[\mathbf{H}]_{n,k} = H_n(\mathbf{m}_k)$. A block diagram of the standard BSS system is shown in Fig. 1.

The aim of any BSS algorithm is to design the unmixing matrix $\mathbf{H}$ such that y_1 and y_2 are “separated”. Typical criteria to achieve this are minimization of mutual information, entropy maximization, maximum likelihood and central limit theorem. A general overview of recent BSSD methods can be found in [5].

2.2. Performance Measure

Let $g_1(\mathbf{s})$ denote the overall system transfer function from the source located at $\mathbf{s}$ to the first output, and similarly for $g_2(\mathbf{s})$ and the second output. The matrix of system transfer functions is:

$$\mathbf{G} = \mathbf{H}\mathbf{A}, \quad (5)$$

where the (n, k) th element is $[\mathbf{G}]_{n,k} = g_n(\mathbf{s}_k)$.

Our aim in this paper is to study the influence of reverberation on the performance of BSS algorithms. In particular we wish to develop expressions for the system transfer functions that one could expect to obtain in a reverberant room, when the unmixing matrix is designed to satisfy certain criteria. Towards this end, and without loss of generality, we will examine the system transfer functions from each source to the first output y_1 . Because of the symmetry of the system, analogous results will hold for the second output, y_2 , which we therefore do not show here. We assume that the permutation problem inherent to frequency-domain BSS algorithms has been solved, and that the first output signal corresponds to the first source signal, and similarly for the second output and the second source signal.

Let $g(\mathbf{s}_n) \equiv g_1(\mathbf{s}_n)$ describe the system TFs from the sources to the first output, and let $H_k \equiv H_1(\mathbf{m}_k)$. Also, let $\mathbf{g} \triangleq [g(\mathbf{s}_1), g(\mathbf{s}_2)]^T$, and $\mathbf{h} \triangleq [H_1, H_2]^T$, so (5) becomes $\mathbf{g}^T = \mathbf{h}^T \mathbf{A}$. In separating the two signals, the BSSD algorithm must “undo” the effect of the ATF to some extent. Suppose that the BSSD algorithm operates such that only the direct-path components of $\mathbf{A}$ are unmixed.¹ De-

¹BSSD algorithms that operate essentially as null beamformers (e.g. [4]) will have this property.

note the resulting demixing vector as $\mathbf{h}_d$, which satisfies $\hat{\mathbf{g}} = \mathbf{A}_d^T \mathbf{h}_d$ where $\mathbf{A}_d$ is a matrix containing only the direct-path components of $\mathbf{A}$ (the (n, k) th element is given by $[\mathbf{A}_d]_{n,k} = A_d(\mathbf{s}_k, \mathbf{m}_n)$), and $\hat{\mathbf{g}}$ are the desired unmixed system TFs (note that $\hat{\mathbf{g}} = [1, 0]^T$ would imply perfect separation of the direct paths). Because only the direct paths are separated, the overall system TFs $\mathbf{g}$ will not match $\hat{\mathbf{g}}$ precisely.

Suppose now the first source moves to a new location $\hat{\mathbf{s}}_1$ with $\|\mathbf{s}_1 - \hat{\mathbf{s}}_1\| = \Delta$, *before* the unmixing filters are able to re-converge. The signals will now be mixed according to a different matrix $\mathbf{A}_\Delta$ with the new subscript denoting that the mixing process is now a function of the source displacement Δ . The actual system transfer function will then be $\mathbf{g} = \mathbf{A}_\Delta^T \mathbf{h}_d$. Again, the overall system TFs will differ from the desired TFs.

The questions we wish to address are: (i) What is the effect on the first source signal at the first output; and (ii) How much of the second source signal remains at the first output. The first question can be answered by considering the *deviation term*: $E\{|g(\hat{\mathbf{s}}_1) - \hat{g}(\mathbf{s}_1)|^2\}$. Similarly, the second question can be answered by considering the *residual term*: $E\{|g(\hat{\mathbf{s}}_2) - \hat{g}(\mathbf{s}_2)|^2\}$. These are the performance measures that we will use to analyze the expected performance of BSS algorithms, and in the sequel we derive closed-form expressions for these terms under various conditions.

3. RESULTS

3.1. Separation of Direct Paths

Theorem 1 *If the unmixing matrix is designed to separate the direct-path components of the mixed signals, the deviation of the first source signal at the first output is*

$$E\{|g(\mathbf{s}_1) - \hat{g}(\mathbf{s}_1)|^2\} = \varepsilon_R, \quad (6)$$

and the residual of the second source at the first output is

$$E\{|g(\mathbf{s}_2) - \hat{g}(\mathbf{s}_2)|^2\} = \varepsilon_R, \quad (7)$$

where

$$\varepsilon_R = \left(\frac{1 - \alpha}{\pi S \alpha} \right) \left(|H_1|^2 + |H_2|^2 + 2\Re\{H_1 H_2^*\} \frac{\sin(k\|\mathbf{m}_1 - \mathbf{m}_2\|)}{k\|\mathbf{m}_1 - \mathbf{m}_2\|} \right) \quad (8)$$

where $\Re\{\cdot\}$ is the real component of the argument, S is the total surface area of the walls and α is the average absorption coefficient of the walls.

This theorem can be proved using results from [3]. Note that it also describes the deviation and residual with the sources in their nominal positions.

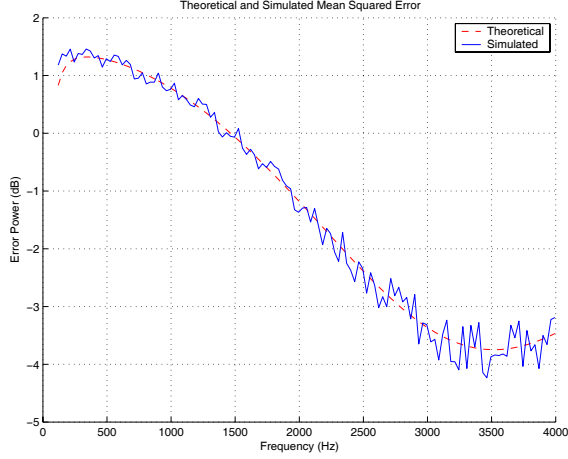


Fig. 2. Comparison of experimental and theoretical curves of (6).

3.2. Separation of Direct Paths, Moving Sources

Theorem 2 Assume the unmixing matrix is designed to separate the direct-path components of the mixed signals with the source signals in default positions $\mathbf{s}_1$ and $\mathbf{s}_2$. If the first source moves to a position $\hat{\mathbf{s}}_1$ that is a distance Δ from $\mathbf{s}_1$ in an arbitrary direction, the deviation of the first source at the first output is

$$E\{|g(\hat{\mathbf{s}}_1) - \hat{g}(\mathbf{s}_1)|^2\} = \varepsilon_{\Delta}(\mathbf{s}_1) + \varepsilon_R, \quad (9)$$

where ε_R is given by (8), and $\varepsilon_{\Delta}(\mathbf{s}_1) \equiv \varepsilon_{\Delta}(\mathbf{s} = \mathbf{s}_1)$ with

$$\begin{aligned} \varepsilon_{\Delta}(\mathbf{s}) = & |H_1|^2 \left(|A_d(\mathbf{s}, \mathbf{m}_1)|^2 - 2 |A_d(\mathbf{s}, \mathbf{m}_1)|^2 \frac{\sin(k\Delta)}{k\Delta} \right. \\ & + \left. \frac{1}{2(4\pi)^2 \Delta \|\mathbf{s} - \mathbf{m}_1\|} \ln \left| \frac{\|\mathbf{s} - \mathbf{m}_1\| + \Delta}{\|\mathbf{s} - \mathbf{m}_1\| - \Delta} \right| \right) \\ & + |H_2|^2 \left(|A_d(\mathbf{s}, \mathbf{m}_2)|^2 - 2 |A_d(\mathbf{s}, \mathbf{m}_2)|^2 \frac{\sin(k\Delta)}{k\Delta} \right. \\ & + \left. \frac{1}{2(4\pi)^2 \Delta \|\mathbf{s} - \mathbf{m}_2\|} \ln \left| \frac{\|\mathbf{s} - \mathbf{m}_2\| + \Delta}{\|\mathbf{s} - \mathbf{m}_2\| - \Delta} \right| \right) \\ & + 2\Re \left\{ H_1 H_2^* A_d(\mathbf{s}, \mathbf{m}_1) A_d^*(\mathbf{s}, \mathbf{m}_2) \left(1 - 2 \frac{\sin(k\Delta)}{k\Delta} \right. \right. \\ & \left. \left. + e^{jk\Delta^2 \mu_s} \frac{\sin(k\zeta_s \Delta)}{k\zeta_s \Delta} \right) \right\} \end{aligned} \quad (10)$$

where

$$\begin{aligned} \zeta_s = & \frac{1}{\|\mathbf{s} - \mathbf{m}_2\| \|\mathbf{s} - \mathbf{m}_1\|} \times \\ & [(\|\mathbf{s} - \mathbf{m}_2\| \mathcal{X}_1 - \|\mathbf{s} - \mathbf{m}_1\| \mathcal{X}_2)^2 \\ & + (\|\mathbf{s} - \mathbf{m}_2\| \mathcal{Y}_1 - \|\mathbf{s} - \mathbf{m}_1\| \mathcal{Y}_2)^2 \\ & + (\|\mathbf{s} - \mathbf{m}_2\| \mathcal{Z}_1 - \|\mathbf{s} - \mathbf{m}_1\| \mathcal{Z}_2)^2]^{1/2}, \end{aligned} \quad (11)$$

$$\mu_s = \frac{1}{2} \left(\frac{1}{\|\mathbf{s} - \mathbf{m}_1\|} - \frac{1}{\|\mathbf{s} - \mathbf{m}_2\|} \right). \quad (12)$$

Similarly, if the second source moves to a position $\hat{\mathbf{s}}_2$ that is a distance Δ' from $\mathbf{s}_2$ in an arbitrary direction, the residual of the second source at the first output is

$$E\{|g(\hat{\mathbf{s}}_2) - \hat{g}(\mathbf{s}_2)|^2\} = \varepsilon_{\Delta'}(\mathbf{s}_2) + \varepsilon_R, \quad (13)$$

where ε_R is given by (8), and $\varepsilon_{\Delta'}(\mathbf{s}_2)$ is given by (10) with $\mathbf{s} = \mathbf{s}_2$ and $\Delta = \Delta'$.

This theorem can again be proved using results from [3]. Note that $\varepsilon_{\Delta}(\mathbf{s}) \rightarrow 0$ as $\Delta \rightarrow 0$. Observe that both (6) and (9) are factored into two separate components: one that describes the effect of movement, and the other that describes the effects of reverberation. It is thus of interest to investigate their relative impact on BSSD performance.

3.3. Simulations and Discussion

To verify these theoretical expressions we check them against simulations using the *Image Model* [1]. We consider a room of dimensions [6.4, 5, 4] metres, with a reverberation time T_{60} of 0.45s. The relative geometry of the system has the sources 0.5m apart, parallel to the receivers line-array which are 0.1m apart. The mid-point of the array is 1m from the corresponding mid-point between the sources. We determine the frequency responses between the sources and each one of the receivers, and average the results over 100 *Monte-Carlo* simulations where the whole setup is randomly rotated and translated within the room. The values of the elements of $\mathbf{h}_d$ were calculated such that $[1, 0]^T = \mathbf{A}_d^T \mathbf{h}_d$ is satisfied, i.e. direct paths are perfectly demixed. The resulting simulated plots show good agreement with the respective theoretical expressions. This can be seen in Fig. 2 and Fig. 3 for Eq. (6) and Eq. (9) respectively.

The error level predicted by (9) is a function of two terms: $\varepsilon_{\Delta}(\mathbf{s})$ and ε_R . We want to examine their relative effect, i.e. the error level due to displacement and due to reverberation. Fig. 4 shows ε_R for different reverberation times. Fig. 5 shows the added error due to displacement $\varepsilon_{\Delta}(\mathbf{s})$. Comparison of the two figures shows that reverberation is a significant contributor to the error level, even for low values of T_{60} . It appears that for direct-path unmixing, movement only becomes significant when larger than 5cm. It can also be seen that the reverberant error is dominant at low frequencies while the one due to displacement dominates at higher frequencies.

The described analysis only refers to direct-path unmixing. It is of interest to examine the two error terms when exact (direct and reverberant paths) unmixing is performed. This was done through simulations. The results shown in Fig. 6 compare directly the exact-path system with the respective unmixing of direct-path. It appears that for the chosen geometry, exact-path equalization is more robust only

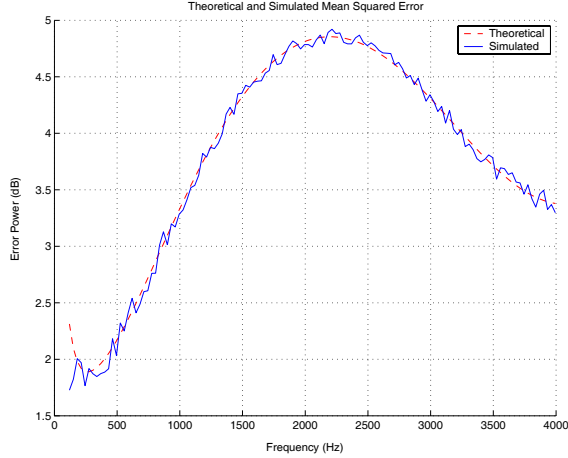


Fig. 3. Comparison of experimental & theoretical curves of (9). The first source is displaced by $\Delta = 5\text{cm}$.

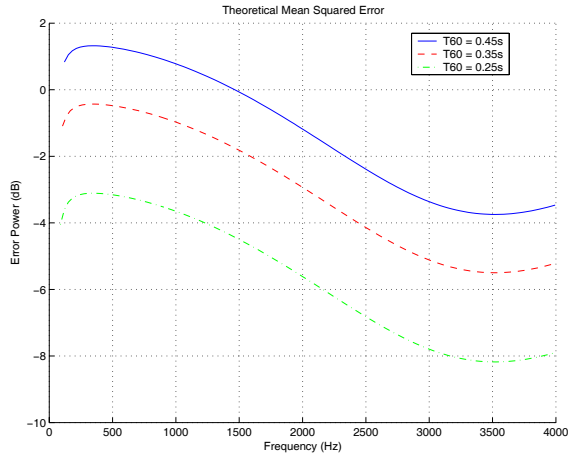


Fig. 4. Variation of ε_R in (8) for changing T_{60} .

for displacements smaller than a few cm. In a sense, this suggests that the choice of the unmixing scheme depends on the application and the environment variations expected. Nevertheless BSSD algorithms that are constrained to unmix only direct-path components seem to have more predictable behaviour in reverberant rooms. Their computational complexity is significantly smaller as well.

4. CONCLUSIONS

We have investigated the effect of source movement and reverberation on BSSD algorithms. We found that when the movement is larger than a few centimeters the robustness of the algorithms attempting to unmix only the direct-path components of the ATF match and even outperform BSSD algorithms that attempt to unmix the entire reverberant ATF. This suggests that BSSD algorithms that are constrained to unmix only direct-path components will be more robust to

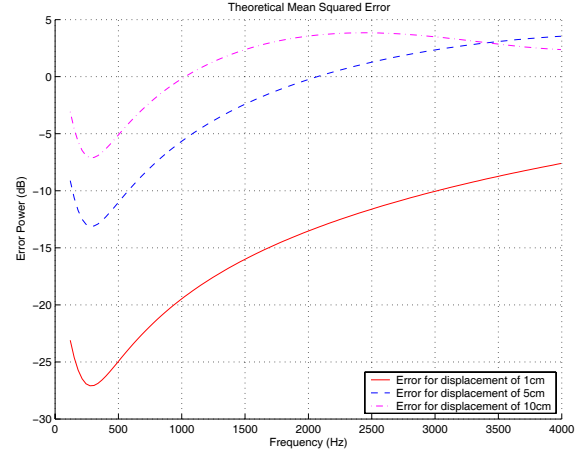


Fig. 5. Variation of $\varepsilon_{\Delta}(s_1)$ in (10) for different displacements Δ .

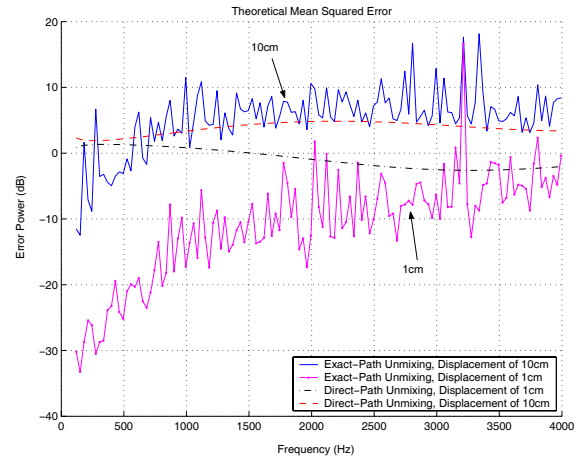


Fig. 6. Comparison of exact and direct-path unmixing for different displacements Δ .

source movement in a real application.

5. REFERENCES

- [1] J.B. Allen, D.A. Berkley, "Image method for efficiently simulating small-room acoustics", *J. Acoust Soc. Amer.*, vol. 65, no. 4, pp. 943-950, 1979.
- [2] A. Koutras, E. Dermatas, G. Kokkinakis, "Blind Speech Separation Of Moving Speakers In Real Reverberant Environments", *Proc. ICASSP 2000*, vol. 2, pp.1133-1136, 2000.
- [3] F. Talantzis, D.B. Ward, "Robustness Of Multi-Channel Equalization In Acoustical Reverberant Environments", *J. Acoust Soc. Amer.*, vol.114, no.2, pp.833-841, 2003.
- [4] J.P. Rosca, J. Ruanaidh, A. Jourjine, S. Rickard "Broad-band Direction-Of-Arrival Estimation Based on Second Order Statistics" *NIPS 1999*, pp.775-781, 1999.
- [5] A. Cichocki, S. Amari, "Adaptive Blind Signal and Image Processing", *John Wiley & Sons*, 2002.