# CRITERION FOR SENSOR PLACEMENT IN ROOM-BASED NEARFIELD MICROPHONE ARRAYS

*Darren B. Ward*

Department of Electrical and Electronic Engineering
Imperial College of Science, Technology and Medicine
Exhibition Road, London SW7 2BT, U.K.

## ABSTRACT

The choice of sensor positions plays a significant role in determining the performance of a nearfield microphone array. In this paper we propose a criterion for sensor placement that attempts to minimize the effect of reverberation for a source within a reverberant room. Rather than assuming a single source position, the criterion attempts to provide the best performance taking into account the probability density function of the source position. One result of applying the proposed method is that for speech acquisition in a room, it may be better to distribute the microphones throughout the room rather than placing them together in a single array (as conventional farfield array theory would dictate).

## 1. INTRODUCTION

There has been much work on microphone arrays over the past 20 or so years [1]. Traditionally, microphone array designs were based on classical array theory, developed for farfield sources with narrow frequency bands. Such classical theory suggests that sensors should be placed with half-wavelength spacing in order to avoid spatial aliasing. Since speech is broadband in nature, microphone arrays have typically used a geometry based on harmonic nesting, in which the array is composed of a set of nested equispaced arrays, with each subarray being designed for a particular frequency range. These considerations are all valid where a farfield source is assumed.

When the desired source is in the nearfield, however, it has been shown that the optimum weights (that maximize the array gain) are effectively inversely proportional to the distance from the sensor to the source [4]. In this case the best performance will be produced by having sensors close to the desired source. Thus, in this paper we propose a new criterion for sensor placement in microphone arrays which is based on maximizing the array gain by placing the sensors according to the probability density function (pdf) describing the expected location of the desired source. Other criteria for sensor placement in microphone arrays have been considered in [2].

## 2. SIGNAL MODEL AND FORMULATION

Consider an array of $N$ microphones. Assume that the desired source is a point source radiator, and the direct-path acoustic transfer function (TF) from the source to the $n$th microphone output is

$$a_n(\omega) = \frac{1}{d_n} e^{-j\omega c^{-1} d_n}, \quad n = 1, \dots, N, \tag{1}$$

where $\omega = 2\pi f$ is the angular frequency, $c$ is the speed of wave propagation, and $d_n$ is the distance from the source to the $n$th microphone which is given by

$$d_n = ||\mathbf{p}_s - \mathbf{p}_n|| \tag{2}$$

where $\mathbf{p}_s = [x_s, y_s, z_s]^T$ is the source position vector, $\mathbf{p}_n = [x_n, y_n, z_n]^T$ is the position vector of the $n$th microphone, and $||\cdot||$ is the vector 2-norm.[1]

At a given frequency $\omega$, the vector of signals received at the array is given by

$$\mathbf{x}(\omega) = \mathbf{a}(\omega)s(\omega) + \mathbf{v}(\omega), \tag{3}$$

where

$$\mathbf{a}(\omega) = [a_1(\omega), \dots, a_N(\omega)]^T \tag{4}$$

is the vector of direct-path TFs, $s(\omega)$ is the desired source signal, and $\mathbf{v}(\omega)$ is the vector of noise signals obtained at the array. In this paper we only consider the case where a single source is used in a reverberant environment. Thus, the noise vector $\mathbf{v}(\omega)$ contains effects due to reverberation alone. To simplify notation we drop the explicit dependence on frequency $\omega$ in the remainder.

Define the noise covariance matrix as

$$\mathbf{Q} = E\{\mathbf{v}\mathbf{v}^H\}. \tag{5}$$

We will model the reverberation as being due to a diffuse noise field, so the $(m, n)$th element of the noise covariance matrix is given by the well known expression [3]

$$\mathbf{Q}_{(m,n)} = \text{sinc}(\omega c^{-1} ||\mathbf{p}_n - \mathbf{p}_m||), \tag{6}$$

[1] This formulation assumes omni-directional microphones. Directional microphones can be easily incorporated by including a suitable directional term in (1).

where $\text{sinc}(x) = \sin x/x$. This diffuse model for reverberation becomes valid for frequencies above the Schroeder frequency, $f_{\text{Sch}} = 2000(T_{60}/V)^{1/2}$, where $V$ is the volume of the room and $T_{60}$ is the reverberation time (defined as the time taken for the sound pressure level to decay by 60 dB once the source has stopped).

Assume that a set of weights $\mathbf{h}$ is applied to the received array signals to form the signal estimate

$$\hat{s} = \mathbf{h}^H \mathbf{x}, \tag{7}$$

where $^H$ denotes Hermitian (complex conjugate) transpose. There are several optimization criteria that have been proposed for design of the weight vector.

One common criterion is to minimize the output noise power of the array subject to a linear constraint on the response of the array to to the desired signal, i.e.,

$$\min_{\mathbf{h}} \mathbf{h}^H \mathbf{Q} \mathbf{h} \;\text{ subject to } \mathbf{h}^H \mathbf{a} = r, \tag{8}$$

where $\mathbf{a}$ is the direct-path TF vector for the source location (assumed known), and $r$ is the desired response to the source signal (typically chosen as a suitable delay). The solution to this problem is

$$\mathbf{h}_{\text{o}} = \frac{\mathbf{Q}^{-1}\mathbf{a} r^*}{\mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a}}, \tag{9}$$

where $^*$ denotes complex conjugate.

These weights also solve the problem of maximizing the array gain, where array gain is defined as

$$G = \frac{\mathbf{h}^H \mathbf{a}\mathbf{a}^H \mathbf{h}}{\mathbf{h}^H \mathbf{Q} \mathbf{h}}. \tag{10}$$

This was the problem considered in [4]. The value of $\mathbf{h}$ that maximizes this ratio of two quadratic forms is

$$\mathbf{h}_{\text{o}} = \mathbf{Q}^{-1}\mathbf{a}\;\alpha, \tag{11}$$

where $\alpha$ is a scalar constant. Choosing $\alpha = r^*/(\mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a})$ will impose the above linear constraint on the source response.

With this optimum value of $\mathbf{h}_{\text{o}}$, the output noise power is

$$J = \frac{|r|^2}{\mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a}}, \tag{12}$$

where $|r|^2 = 1$ if $r$ is chosen as a pure delay. In the remainder we will assume that $|r|^2 = 1$.

## 3. SENSOR PLACEMENT CRITERION

Observe from (1) and (6) that the output noise power (12) will depend explicitly on the sensor positions, and one could therefore choose the sensor positions to minimize the output noise power. For a single source position, this would be found by minimizing (12).

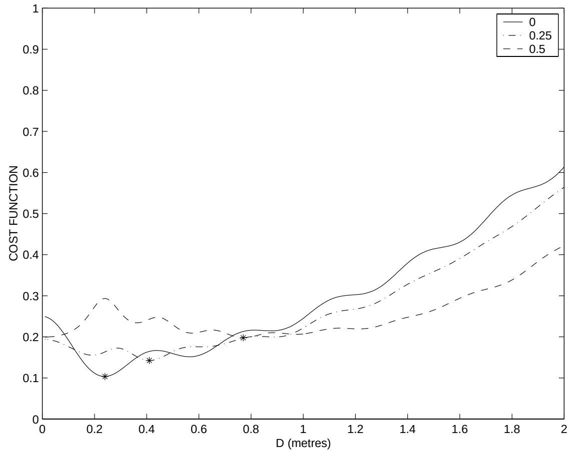


Figure 1: Variation of output noise power (12) versus intersensor spacing $D$, for three different positions of the source signal ($x_s = 0, 0.25, 0.5$ m).

As an example, consider an array with two microphones, located at $\mathbf{p}_1 = [-D/2, 0, 0]^T$ and $\mathbf{p}_2 = [D/2, 0, 0]^T$, respectively. Assume the desired source is located at $\mathbf{p}_s = [x_s, 0.5, 0]^T$. At a frequency of 1000 Hz, the variation of the cost function (12) with $D$ for different values of $x_s$ is shown in Fig. 1. For each value of $x_s$, the optimum spacing $D_{\text{o}}$ is marked by an asterisk. Notice that the optimum spacing varies significantly for different source positions.

In practice one will never know *a priori* where the source is located, and in fact, the source will typically be moving within some region. Therefore, in this paper we propose to choose the sensor positions to minimize the expected output noise power, where expectation is taken with respect to the distribution of the source position. In the remainder we assume that the probability density function (pdf) of the source location is known. This is less restrictive than designing the array assuming that the source position is fixed. For example, one could design the array to give reasonable performance for a source at any position within a specified region; in this case the pdf would be chosen as a uniform distribution. In other cases (for example, a microphone array placed on a computer monitor where the user will typically be located directly in front of the screen but may move slightly away from this position) a Gaussian pdf may be more appropriate.

Assume that the source location vector $\mathbf{p}_s$ has a known probability density function $g(\mathbf{p}_s)$. Using (12), the expected noise power is given by

$$E\{J\} = \int_{-\infty}^{\infty} \frac{1}{\mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a}}\, g(\mathbf{p}_s)\; d\mathbf{p}_s. \tag{13}$$

The proposed criterion is to choose the set of sensor locations to minimize the expected output noise power:

$$\{\mathbf{p}_n\} = \arg\min_{\mathbf{p}_n} \int_{-\infty}^{\infty} \frac{1}{\mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a}}\, g(\mathbf{p}_s)\; d\mathbf{p}_s. \tag{14}$$

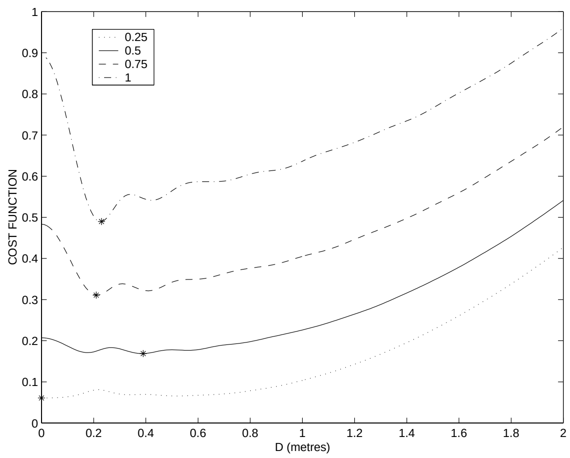


Figure 2: Variation of (16) versus intersensor spacing $D$ at a frequency of 1000 Hz for a two-element array, with four different source distances ($y_s = 0.25, 0.5, 0.75, 1$ m).

Even for the simple two-sensor example considered above, the problem of finding the sensor positions $\mathbf{p}_n$ that solve (14) is nonlinear, and finding a closed-form solution appears to be nontrivial. We will therefore consider numerical solutions in the remainder of this paper.

Returning to the two-sensor example considered in Fig. 1, assume that the source is positioned along a line parallel to the array at a distance of $y_s = 0.5$ m, and that the source position is uniformly distributed within a region of $x_s \in [-L/2, L/2]$. The source location pdf is therefore

$$g(\mathbf{p}_s) = \begin{cases} \frac{1}{L}, & -L/2 \leq x_s \leq L/2 \\ 0 & \text{otherwise}, \end{cases} \tag{15}$$

and we have

$$E\{J\} = \frac{1}{L} \int_{-L/2}^{L/2} J \; dx_s. \tag{16}$$

A plot of the cost function (16) versus spacing $D$, for $L = 1$ m and at a frequency of 1000 Hz is shown by the solid line in Fig. 2.[2] The optimum spacing (marked by an asterisk) is given by $D_o = 0.39$ m.

Clearly, the optimum spacing also depends on the distance from the source to the array. To investigate this, the resulting values of (16) for different source distances are shown in Fig. 2. As one would expect, the further the source is away from the array the worse the noise performance. However, notice also from Fig. 2 that for a source distance of 0.25 m the optimum sensor spacing is $D = 0$ m.[3] In other words, in this case it appears to be better to have only a single sensor rather than two! This result seems surprising.

[2] The integral in (16) was calculated by summing over 21 equally spaced source locations within the region $x_s \in [-L/2, L/2]$.

[3] In fact, the minimum was $D = 1 \times 10^{-6}$ m, since $\mathbf{Q}$ is singular if $D = 0$.

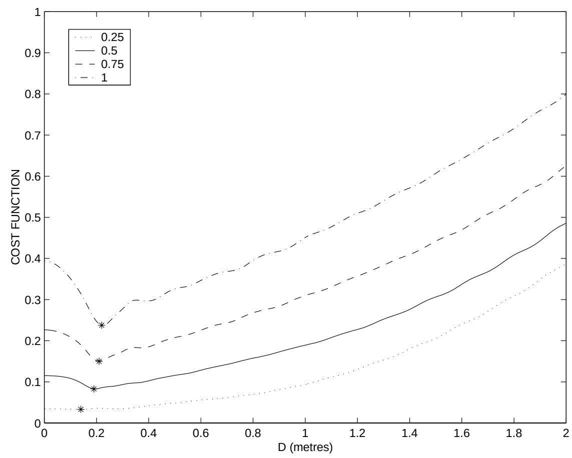


Figure 3: As in Fig. 2, but using a four-element array.

This experiment was repeated using a four-element array. Again the sensors were equally spaced, and the source was positioned along the same line parallel to the array. The variation of the cost function (16) as a function of intersensor spacing $D$, for $L = 1$ m and a frequency of 1000 Hz, is shown in Fig. 3 for different source distances. As one would expect, increasing the number of sensors improves the noise performance (i.e., reduces the value of $E\{J\}$). In comparing Figs. 2 and 3, notice that for the optimum spacing the value of $E\{J\}$ is approximately proportional to $1/N$ (where $N$ is the number of sensors). Simulations with different numbers of equispaced sensors confirm that this relation holds approximately. However, this is not true for all values of $D$, and in fact, for large values of $D$ increasing the number of sensors provides very little improvement in the expected output noise power.

There is one other comment to make regarding these results. Although in each case an optimum value of $D$ can clearly be found by minimizing (16), one might ask whether this is indeed necessary. Referring to Figs. 2 and 3, one notices that a spacing of a half-wavelength (or 0.17 m) results in a value of $E\{J\}$ that is negligibly different from that obtained using the optimum spacing. It therefore appears that even for nearfield arrays, a spacing of a half-wavelength is a reasonably good choice for a linear array with equally spaced sensors designed for operation at a single frequency.

## 4. BROADBAND DESIGN

The formulation considered so far has considered operation at a single frequency. However, microphone arrays designed for speech pickup must cover a wide frequency band of several octaves. In this section we extend the previous formulation to a broadband design and consider an example for speech pickup in a room.

For a broadband design, the proposed criterion for sen-

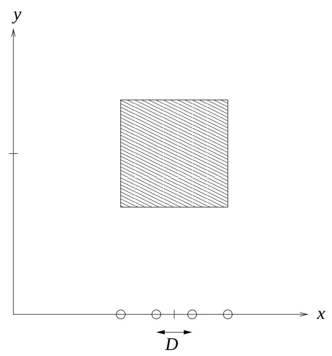


Figure 4: Geometry for broadband example. Source locations are indicated by the shaded region, microphone location are denoted by the circles.

sor placement can be modified to yield:

$$\{\mathbf{p}_n\} = \arg\min_{\mathbf{p}_n} \int_{\omega_L}^{\omega_U} \int_{-\infty}^{\infty} \frac{g(\mathbf{p}_s)}{\mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a}} \, d\mathbf{p}_s \, d\omega, \qquad (17)$$

where $\omega_L$ and $\omega_U$, respectively, are the lower and upper band edges of the desired frequency range.

We now consider an example to demonstrate application of the broadband criterion (17). Consider an array of four microphones, equally spaced with a spacing of $D$ m. The source position is uniformly distributed within a square region at the same height as the array, and the center of the source distribution is equidistant from the $x$ and $y$ axes and is collinear with the center of the array. Refer to Fig. 4 for the geometry. The frequency band of interest was chosen as 300 Hz to 3400 Hz, and the integration over frequency in (17) was performed by summing over the band in intervals of 100 Hz.

Results of applying (17) to find the optimum spacing $D$ are shown in Table 1(a), in which $D_{\mathrm{o}}$ is the optimum spacing and $J_{\mathrm{o}}$ is the corresponding minimum value of the double integral in (17). Two different sizes of the source region are considered, 1 m $\times$ 1 m and 2 m $\times$ 2 m. For each case, four different positions of the source region are considered, where the $(x, y)$ location of the center of the source region is given in the left hand column. In each example, the array is positioned along the $x$-axis with its center collinear with the center of the source region. As one would expect, performance improves (i.e., $J_{\mathrm{o}}$ decreases) as the source region moves closer to the array.

We also consider an alternative array geometry with two separate arrays, each containing two sensors separated by a distance $D$, where one array is located along the $x$-axis and the other along the $y$-axis of Fig. 4. The center of each array is collinear with the center of the source region. Results are shown in Table 1(b). In comparing Table 1(a) and (b), note that in each case the two perpendicular 2-element arrays perform better than the single 4-element array, with normalized improvements ranging from 3% to 20%.

| Center of source region | 1 m $\times$ 1m | | 2 m $\times$ 2 m | |
|---|---|---|---|---|
| | $J_{\mathrm{o}}$ | $D_{\mathrm{o}}$ | $J_{\mathrm{o}}$ | $D_{\mathrm{o}}$ |
| (3.0,3.0) | 2.32 | .38 | 2.47 | .36 |
| (2.5,2.5) | 1.64 | .32 | 1.78 | .34 |
| (2.0,2.0) | 1.07 | .28 | 1.22 | .32 |
| (1.5,1.5) | 0.63 | .24 | 0.77 | .32 |

(a) 4-element array

| Center of source region | 1 m $\times$ 1m | | 2 m $\times$ 2 m | |
|---|---|---|---|---|
| | $J_{\mathrm{o}}$ | $D_{\mathrm{o}}$ | $J_{\mathrm{o}}$ | $D_{\mathrm{o}}$ |
| (3.0,3.0) | 2.25 | .52 | 2.26 | .52 |
| (2.5,2.5) | 1.57 | .50 | 1.59 | .48 |
| (2.0,2.0) | 1.01 | .46 | 1.04 | .46 |
| (1.5,1.5) | 0.57 | .40 | 0.61 | .48 |

(b) Two perpendicular 2-element arrays

Table 1: Broadband design example (refer to text for details).

## 5. CONCLUSIONS

A new criterion for sensor placement in microphone arrays has been proposed in which the microphones are located to minimize the expected output noise power. Rather than assuming the source is in a fixed position, the source is assumed to be located within a certain region with a given probability density function. One new result is that for speech acquisition in a room, it may be better to distribute the sensors throughout the room rather than placing them together in a conventional array geometry. Finally, it must be pointed out that the present investigation is preliminary and has been based on numerical simulations. Further investigation is required to enable one to draw more general theoretical conclusions.

## 6. REFERENCES


[1] M.S. Brandstein and D.B. Ward, editors. *Microphone Arrays: Signal Processing Techniques and Applications*. Springer-Verlag, Berlin, 2001.

[2] S. Gazor and Y. Grenier. Criteria for positioning of sensors for a microphone array. *IEEE Trans. Speech Audio Processing*, 3(4):294–303, July 1995.

[3] P.A. Nelson and S.J. Elliot. *Active Control of Sound*. Academic Press, London, 1992.

[4] J.G. Ryan and R.A. Goubran. Array optimization applied in the near field of a microphone array. *IEEE Trans. Speech Audio Processing*, 8(2):173–176, March 2000.