BEAMFORMING FOR A SOURCE LOCATED IN THE INTERIOR OF A SENSOR ARRAY

Darren B. Ward

School of Electrical Engineering, ADFA
University College, University of New South Wales
Canberra ACT 2600
d-ward@ee.adfa.edu.au

Robert C. Williamson

Department of Engineering, FEIT
The Australian National University
Canberra ACT 0200
bob.williamson@anu.edu.au

ABSTRACT

We introduce a framework for acquiring a signal from a source that is located within the midst of a randomly distributed sensor array. This problem arises in speech acquisition with microphone arrays. Based on psychoacoustic considerations, we formulate a constrained optimization problem in which the array weights are chosen to minimize the response to farfield sources while maintaining a unity-gain constraint for the source signal.

1. INTRODUCTION

Over the past few decades there have been many developments in the field of array signal processing. However, the vast majority of array theory developed to date has been for farfield sources and a linear array geometry (or some other regular equally-spaced pattern). New theory has recently been developed that is suitable for nearfield sources [1], although techniques that remove the limitations of farfield sources and an equally-spaced array geometry tend to be the exception.

In this paper we introduce a framework for dealing with a general beamforming problem in which the array sensors may be placed in arbitrary (but known) locations, and the source is located within the convex hull spanned by the array. We will refer to this as *interior field* beamforming.

A related problem was recently studied in [2], in which the authors considered randomly distributed sensors whose positions were unknown. They presented a blind beamforming technique in which the array weights were chosen to maximize the output power. Our approach differs in that we assume the sensor locations are known, and we choose the array weights to improve the intelligibility of an acquired speech signal.

Our motivation for studying the interior field problem is room-based hands-free speech acquisition where an array of microphones is located randomly on the walls of a room, and the aim is to obtain the best possible reproduction of a talker's speech signal. Although this is a broadband signal acquisition problem (since speech is a broadband signal that covers several octaves), for simplicity we will only consider a narrowband formulation in this paper; extension to a more general broadband signal environment will be covered in future work.

This work was supported by the Australian Research Council.

2. THE INTERIOR FIELD PROBLEM

Consider an array of N identical omni-directional sensors located at points $\mathbf{p}_n = [x_n, y_n, z_n]^T \in \mathbb{R}^3, n = 1, \ldots, N$. Apart from assuming the sensor locations are known and unique, we place no other restrictions on these locations. This differs markedly from the vast majority of array literature in which the sensors are placed in a regular pattern, usually a linear equally-spaced geometry.

Assume that a single source is present at the location $\mathbf{s} \in \mathbb{R}^3$ (with $\mathbf{s} \neq \mathbf{p}_n, \forall n$) and that it produces a signal $S(\omega)$, where $\omega = 2\pi f$ is the angular frequency of operation. Denote the signal received at the nth sensor as $M_n(\omega)$, assume that this sensor signal is filtered by $H_n(\omega)$, and that all filter outputs are summed to form the beamformer output:

$$Y(\omega) = \sum_{n=1}^{N} M_n(\omega)\, H_n(\omega). \tag{1}$$

Denote the free-field transfer function (TF) between a source located at $\mathbf{s} \in \mathbb{R}^3$ and the nth sensor by

$$A_{\mathbf{s},n}(\omega) = \frac{1}{d(\mathbf{s}, \mathbf{p}_n)}\, e^{-j\omega c^{-1} d(\mathbf{s},\mathbf{p}_n)}, \tag{2}$$

where c is the speed of wave propagation,

$$d(\mathbf{s}, \mathbf{p}_n) = \|\mathbf{s} - \mathbf{p}_n\| \tag{3}$$

is the distance from $\mathbf{s}$ to $\mathbf{p}_n$, and $\|\cdot\|$ denotes the vector 2-norm. Note that this TF models free-field propagation only, i.e., it does not model multipath propagation such as reverberation and it also neglects absorption effects.

The signal received at the nth sensor can now be modelled as

$$M_n(\omega) = A_{\mathbf{s},n}(\omega)\, S(\omega) + V_n(\omega), \tag{4}$$

where $V_n(\omega)$ represents unwanted noise. We consider this noise term in more detail in the following section.

The beamformer output can be written

$$Y(\omega) = S(\omega) B_{\mathbf{s}}(\omega) + V(\omega), \tag{5}$$

where

$$B_{\mathbf{s}}(\omega) = \sum_{n=1}^{N} A_{\mathbf{s},n}(\omega) H_n(\omega) \tag{6}$$

is the TF from the source to the beamformer output, and

$$V(\omega) = \sum_{n=1}^{N} V_n(\omega) H_n(\omega) \tag{7}$$

is the filtered noise component. The beamforming problem is to design the sensor filters $H_n(\omega)$ so that the source signal $S(\omega)$ is passed to the beamformer output with unity gain, and the unwanted noise $V(\omega)$ is minimized. We will formulate this problem more precisely in the following section.

3. DESIGN CRITERIA

3.1. Unity Gain Response to Source Signal

The first design criterion is that the source signal should be passed to the beamformer output with unity gain. In other words, the TF from the source to the beamformer output should be a pure delay, i.e.,

$$B_{\mathbf{s}}(\omega) = \sum_{n=1}^{N} A_{\mathbf{s},n}(\omega) H_n(\omega) = e^{-j\omega c^{-1}\tau_0}, \tag{8}$$

where τ_0 is the desired delay between the source and the beamformer output (usually referred to as the *modelling delay*). This constraint may be written as

$$\mathbf{a}_{\mathbf{s}}^H \, \mathbf{h} = e^{-j\omega\tau_0} \tag{9}$$

where

$$\mathbf{a}_{\mathbf{s}} = [A_{\mathbf{s},1}(\omega), \ldots, A_{\mathbf{s},N}(\omega)]^H \tag{10}$$

and

$$\mathbf{h} = [H_1(\omega), \ldots, H_n(\omega)]^T. \tag{11}$$

To simplify notation we will drop the explicit dependence on ω in the remainder of the paper. Thus below, ω is arbitrary but fixed.

3.2. Minimize Response to Unwanted Noise

The second criterion is to minimize the response of the beamformer to unwanted noise. Referring to (5), this corresponds to minimizing the $V(\omega)$ term. In an adaptive beamformer this could be achieved by minimizing the beamformer output power, using an algorithm such as [3]. However, we will instead consider an alternative formulation that is possible for the case of a source located within a reverberant room. The formulation presented here results in a straightforward closed-form solution.

For a single source located within a reverberant room, the noise term $V(\omega)$ will consist of reverberation only. There are two commonly used models for room reverberation: the first considers reverberation to be composed of incident sound waves propagating from sources distributed uniformly throughout space, i.e., a *diffuse sound field* [4]; the second models room reverberation as being due to image sources whose locations are obtained by repeatedly mirroring the original source in all walls of the room [5].

Using either model, the noise term can be reduced by minimizing the response of the beamformer to sources located at directions other than the source direction. With the vector notation defined in (10) and (11), the response of a beamformer to a source located at $\mathbf{q} \in \mathbb{R}^3$ is

$$B_{\mathbf{q}} = \mathbf{a}_{\mathbf{q}}^H \mathbf{h}. \tag{12}$$

Reducing the noise term is thus equivalent to

$$\min \;\; \mathbf{h}^H \mathbf{Q} \mathbf{h}, \tag{13}$$

where

$$\mathbf{Q} = \int_{\mathbb{R}^3} \mathbf{a}_{\mathbf{q}} \mathbf{W}_{\mathbf{q}} \mathbf{a}_{\mathbf{q}}^H \; d\mathbf{q} \tag{14}$$

and $\mathbf{W}_{\mathbf{q}}$ is a $(N \times N)$ *weighting matrix*. Selection of the appropriate weighting matrix is a critical design step that will be considered in detail in the following section.

3.3. Optimization Problem

The two design criteria specified in (9) and (13) can be combined to yield the following standard constrained optimization problem:

$$\min_{\mathbf{h}} \mathbf{h}^H \mathbf{Q} \mathbf{h} \tag{15}$$

$$\text{subject to } \mathbf{a}_{\mathbf{s}}^H \mathbf{h} = e^{-j\omega\tau_0}, \tag{16}$$

which has a well-known solution

$$\mathbf{h}_{\text{opt}} = \mathbf{Q}^{-1} \mathbf{a}_{\mathbf{s}} \left[\mathbf{a}_{\mathbf{s}} \mathbf{Q}^{-1} \mathbf{a}_{\mathbf{s}}^H\right]^{-1} \; e^{-j\omega\tau_0}. \tag{17}$$

As well as the optimum solution given by (17) (which requires matrix inversion operations that may be computationally expensive), algorithms such as LMS and RLS that converge to this optimum solution may also be used.

4. WEIGHTING MATRIX

In this section we describe some candidates for the weighting matrix. Motivated by results from psychoacoustics, we champion a specific choice for $\mathbf{W}_\mathbf{q}$ that provides a straightforward closed-form solution for $\mathbf{Q}$.

Denote the (m,n)th element of $\mathbf{Q}$ as $\mathbf{Q}^{(m,n)}$. With the elements of $\mathbf{a}_\mathbf{q}$ given by (2) we have

$$\mathbf{Q}^{(m,n)} = \frac{\mathbf{W}_\mathbf{q}^{(m,n)}}{d(\mathbf{q},\mathbf{p}_m)\, d(\mathbf{q},\mathbf{p}_n)}\, e^{-j\omega c^{-1}[d(\mathbf{q},\mathbf{p}_m)-d(\mathbf{q},\mathbf{p}_n)]} \tag{18}$$

where $\mathbf{W}_\mathbf{q}^{(m,n)}$ is the (m,n)th element of $\mathbf{W}_\mathbf{q}$ and $d(\cdot)$ is defined by (3).

4.1. Diffuse Sound Field

If room reverberation is modelled as a diffuse sound field, then we assume the noise term is composed of incident waves emanating from point sources uniformly distributed in $\mathbb{R}^3$. In this case the weighting matrix is given by

$$\mathbf{W}_\mathbf{q}^{(m,n)} = \delta(m-n), \tag{19}$$

and $\mathbf{Q}$ becomes

$$\mathbf{Q} = \int_{\mathbb{R}^3} \mathbf{a}_\mathbf{q}\, \mathbf{a}_\mathbf{q}^H \; d\mathbf{q}. \tag{20}$$

However, there are obvious problems in evaluating $\mathbf{Q}$ explicitly. First, because of the definition of the elements of $\mathbf{a}_\mathbf{q}$ (2) there are singularities at $\mathbf{q} = \mathbf{p}_n, n = 1,\ldots,N$. Second, even without these singularities, the integrand is a non-linear function of $\mathbf{q}$, and no closed-form solution can be found for $\mathbf{Q}$.

4.2. Speech Intelligibility

Our motivation for studying the interior field problem was to obtain a technique for reducing reverberation from a speech signal acquired in a room using an array of microphones. It is therefore of interest to consider how the weighting matrix might be selected to maximize the speech intelligibility of the acquired signal.

Several psychoacoustic studies have been performed to determine the effect of reverberation on speech intelligibility [4]. One of the first of these, undertaken by Thiele in 1953, found that reflections with delay times of greater than 50 ms were detrimental to speech intelligibility. Over the years other studies have been performed to revise this estimate (see [4] for details), although most of these confirm that Thiele's original measure is reasonably accurate.

Let d_{50} denote the distance that sound travels in 50 ms. Using $c = 340$ m/s (the speed of sound propagation in air) gives $d_{50} = 17$ m. With the criterion of reducing the detrimental effects of reverberation on speech intelligibility, we can therefore define the weighting matrix as

$$\mathbf{W}_\mathbf{q}^{(m,n)} = \begin{cases} \delta(m-n), & \|\mathbf{q}-\mathbf{s}\| > d_{50} \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

giving

$$\mathbf{Q} = \int_{\mathbf{q}:\|\mathbf{q}-\mathbf{s}\|>d_{50}} \mathbf{a}_\mathbf{q}\, \mathbf{a}_\mathbf{q}^H \; d\mathbf{q}. \tag{22}$$

Assuming that the source is less than 17m from each sensor (which will occur in all but the largest rooms), then this choice of $\mathbf{W}_\mathbf{q}$ removes the singularities present in the diffuse sound field model. However, it still does not yield a closed-form solution for $\mathbf{Q}$.

4.3. Farfield Interference

For a microphone array located on the walls of a typical room, an image at 17 m from the source corresponds to a *farfield* source at most frequencies of interest. In order to reduce the deleterious effects of reverberation, we may therefore consider minimizing the response of the beamformer to farfield interference. This can be achieved by defining the weighting matrix such that it has unity diagonal elements for sources located in the farfield, and zero otherwise, i.e.,

$$\mathbf{W}_\mathbf{q}^{(m,n)} = \begin{cases} \delta(m-n), & \|\mathbf{q}-\mathbf{s}\| > d_{\text{FF}} \\ 0, & \text{otherwise} \end{cases} \tag{23}$$

where d_{FF} is the distance at which the farfield approximation becomes valid.

For a source located in the farfield of an array, the propagating wavefront is modelled as a planar wave, and the standard farfield TF between the source and the nth element of the array (normalized to ignore attenuation with distance) is given by

$$\widetilde{A}_{\widetilde{\mathbf{q}},n} = e^{-j\omega c^{-1}\widetilde{\mathbf{q}}^T\mathbf{p}_n}, \tag{24}$$

where the $\widetilde{\ }$ symbol is used to identify a farfield quantity, and

$$\widetilde{\mathbf{q}} = [\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta]^T \tag{25}$$

is the location vector for a farfield source located at an elevation of θ and an azimuth of ϕ.

The $\mathbf{Q}$ matrix may now be written

$$\mathbf{Q} = \int \widetilde{\mathbf{a}}_{\widetilde{\mathbf{q}}}\widetilde{\mathbf{a}}_{\widetilde{\mathbf{q}}}^H \; d\widetilde{\mathbf{q}}, \tag{26}$$

where

$$\widetilde{\mathbf{a}}_{\widetilde{\mathbf{q}}} = [\widetilde{A}_{\widetilde{\mathbf{q}},1}, \ldots, \widetilde{A}_{\widetilde{\mathbf{q}},N}]^H. \tag{27}$$

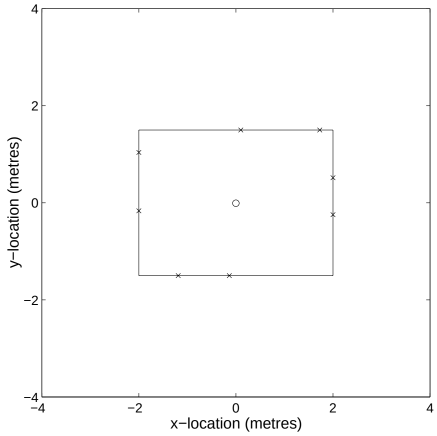

Figure 1: Location of sensors (×) and source (○) within a 3 m × 4 m room.

The elements of $\mathbf{Q}$ are given explicitly by

$$\mathbf{Q}^{(m,n)} = \int e^{-j\omega c^{-1}\widetilde{\mathbf{q}}^T(\mathbf{p}_m - \mathbf{p}_n)}\, d\widetilde{\mathbf{q}}. \qquad (28)$$

Substituting $\widetilde{\mathbf{q}} = [u, v, w]^T$ and carrying out the integration with respect to u, v and w over the unit sphere results in

$$\mathbf{Q}^{(m,n)} = \mathrm{sinc}\left(\tfrac{kx_{mn}}{\pi}\right)\,\mathrm{sinc}\left(\tfrac{ky_{mn}}{\pi}\right)\,\mathrm{sinc}\left(\tfrac{kz_{mn}}{\pi}\right), \qquad (29)$$

where $\mathrm{sinc}(x) = \sin(\pi x)/(\pi x)$, $k = \omega/c$ is the wavenumber, $x_{mn} = x_m - x_n$, $y_{mn} = y_m - y_n$, and $z_{mn} = z_m - z_n$. Hence, the farfield interference weighting matrix defined by (23) results in a straightforward closed-form solution for $\mathbf{Q}$.

5. DESIGN EXAMPLE

Consider an 8-element array placed on the walls of a 3 m × 4 m room as shown in Fig. 1, with a single source present at $\mathbf{s} = [0, 0, 0]^T$. In this simulation we are only considering a two-dimensional environment (and so $z_n = 0$ for all n). The frequency of operation is 500 Hz. The array weights are found from (17) with $\mathbf{Q}$ given by (29). The resulting array response is shown in Fig. 2. In this figure, sources outside the room region represent reverberant images. We note that images located more than about 8 m from the signal source have been attenuated by at least 25 dB.

6. CONCLUSION

Beamforming for a source located in the interior field of a randomly-distributed sensor array is a problem that has been largely ignored in the literature. Such interior field beamforming is motivated by hands-free speech acquisition with a microphone array. In this paper we have introduced this problem and derived a closed-form expression for the optimum array weights that should be used at a single frequency. The optimization problem was formulated as minimizing the response of the beamformer to farfield sources, subject to a unity gain constraint on the response to the desired speech source, and was justified by considering psychoacoustic measures of speech intelligibility. In future work we will extend the technique to a broadband signal environment, test its performance in real rooms, and consider attenuating interfering signals that may also be present in the interior field.

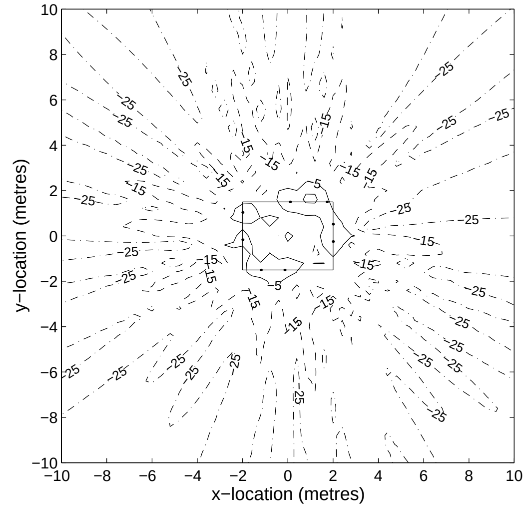

Figure 2: Response of the designed beamformer. The source is located at $\mathbf{s} = [0, 0, 0]^T$, and the sensors are located at the positions denoted by the small dots. The gain at $\mathbf{s}$ is 0 dB.

7. REFERENCES

[1] R.A. Kennedy, D.B. Ward, and T.D. Abhayapala, "Nearfield beamforming using radial reciprocity", *IEEE Trans. Signal Processing*, vol. 47, no. 1, pp. 33–40, Jan. 1999.

[2] K. Yao, R.E. Hudson, C.W. Reed, D. Chen, and F. Lorenzelli, "Blind beamforming on a randomly distributed sensor array system", *IEEE J. Selected Areas Commun.*, vol. 16, no. 8, pp. 1555–1566, Oct. 1998.

[3] O.L. Frost III, "An algorithm for linearly constrained adaptive array processing", *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[4] H. Kuttruff, *Room Acoustics*, Applied Science Publishers, London, 2nd edition, 1979.

[5] J.B. Allen and D.A. Berkley, "Image method for efficiently simulating small-room acoustics", *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.