

Bayesian Direction of Arrival Estimations for Sound Sources Using a Spherical Microphone Array

Weikel, Stephen¹, Xiang, Ning², Landschoot, Christopher ³

Graduate Program in Architectural Acoustics, Rensselaer Polytechnic Institute 110 8th St. Troy, New York 12180, USA

ABSTRACT

A common problem in acoustical applications is the determination of directions of arrival (DoAs) of sound at a receiver. This work aims to address this problem in situations involving potentially multiple simultaneous sound sources by use of spatial filtering, or beamforming, algorithms with a spherical microphone array. This presents a two-level inferential problem of first determining the number of sound sources and then their locations. The solution under investigation here is a probabilistic model-based method, leveraging Bayesian analysis to match analytic models to experimental data. In this method, a large number of models are simulated, predicting the sound field created by various numbers of sources at different positions around the receiver. These predictions are then evaluated against the measured data in order to select the simplest such model that can adequately match observations, thereby estimating first the number of sources, then their DoA information. This paper discusses the impact of microphone array configurations upon DoA resolution. It additionally estimates performance of the method based on experimental results for simultaneous multiple sound sources, particularly in noisy or reverberant environments.

Keywords: Noise Source Localization, Direction of Arrival, Bayesian Analysis, Spherical Harmonics **I-INCE Classification of Subject Number:** 74

1. INTRODUCTION

This work sets out to evaluate a method of determining the directions of arrival (DoAs) of simultaneous sounds from multiple sources at a receiver. This presents a

¹weikes@rpi.edu

²xiangn@rpi.edu

³christopherlandschoot@gmail.com

twofold challenge of determining both the number and location(s) of the detected sound sources. A spherical microphone array is used to collect sound pressure data, from which is extracted directional information via spherical harmonic beamforming. A probabilistic Bayesian algorithm is then employed to determine first the number of discrete sources present, followed by their strengths and locations.

Spherical microphone arrays have seen much development in recent years due to their inherent directional flexibility and ability to support three dimensional beamforming. [1] This latter feature, beamforming, is a process of spatially filtering incoming sound based on its direction of arrival in order to create arbitrary directivity patterns for the array. This is accomplished by applying specific weighting factors to the detected sound pressure around the surface of the spherical array. [2] The array used in this work consists of 16 omnidirictional microphones embedded in a rigid sphere in a nearly-uniform distribution across its surface. Through beamforming, this array can be used to detect and map the directions of incident sound energy.

The challenge then remains to make sense of this energy mapping and to discern from it the number and positions of sound sources outputing the energy. This work investigates the capabilities of a two-tiered model based Bayesian inference algorithm to determine these features. [3] Similar work has been conducted, using two levels of Bayesian inference to determine source number and position, with data collected using different microphone arrays including coprime arrays and a simple two-microphone array. [4] [5] However, to the best of the authors' knowledge, the work described herein is the first time that an algorithm of this sort has been used to analyze data collected with a spherical array.

This paper will present in section 2 a brief description of the beamforming techniques employed, and the formulation of the models to be evaluated. A more thorough discussion of the work will be available in an upcoming paper. [6] Section 3 will present the two-tiered Bayesian framework. Section 4 presents some initial results of this work, while section 5 discusses them and concludes the paper.

2. BEAMFORMING AND MODEL FORMULATION

The technique of beamforming underlies two essential tasks in this work. It is used both to map the incoming sound energy over the sphere, and to formulate models for evaluation by the Bayesian algorithm. For convenience, the relevant calculations are performed in terms of spherical harmonics. These are defined as

$$Y_{n}^{m}(\theta,\phi) = \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_{n}^{m}(\cos\theta) e^{jm\phi},$$
(1)

where θ and ϕ are azimuth and elevation angles and *n* and *m* are, respectively, order and degree of spherical harmonics. [7] $P_n^m(\cdot)$ is the Legendre function of order *n* and degree *m* and $j = \sqrt{-1}$.

In order to express the sound pressure over the sphere in terms of spherical harmonics, the spherical Fourier transform can be used to find the spherical harmonic coefficients p_{nm} by

$$p_{nm}(k,r) = \frac{4\pi}{M} \sum_{i=1}^{M} p(k,r,\theta_i,\phi_i) [Y_n^m(\theta_i,\phi_i)]^*,$$
(2)

where *M* is the number of microphones, θ_i and ϕ_i represent the angular position of the *i*th microphone, *k* is the wave number, *r* is the radius of the sphere, and * denotes the complex conjugate. [2] Spatially filtering pressure data to selectively listen in some direction θ_l, ϕ_l can then be accomplished by applying beamforming weights $w_{nm}^*(k, r, \theta_l, \phi_l)$ to the transformed pressure signal, giving the beamformed data

$$D(\theta_{l},\phi_{l}) = \sum_{n=0}^{N} \sum_{m=-n}^{n} w_{nm}^{*}(k,r,\theta_{l},\phi_{l}) p_{nm}(k,r),$$
(3)

where N is some finite maximum spherical harmonic order and

$$w_{nm}^* = \frac{Y_n^m(\theta_l, \phi_l)}{j^n \mathbf{b}_n(\mathbf{kr})}.$$
(4)

Here, b_n represents modal amplitude which is given by

$$b_n(kr) = j_n - \frac{j'_n(kr)}{h'_n(kr)} h_n(kr),$$
(5)

where j_n and h_n represent, respectively, the spherical Bessel and Hankel functions of the first kind and the prime signifies the derivative with respect to the argument. [2] Sweeping the listening direction of equation 3 over all angles can thus produce a directional map of sound energy over the sphere, with an angular resolution dictated by the maximum order N of the spherical harmonics employed. This maximum order is, in turn, a function of the number and configuration of microphones in the array, with more microphones generally allowing for a higher order spherical harmonic representation of the sound field. [2]

The models to be evaluated by the Bayesian algorithm can be expressed in a similar way. Here, the model needs to represent potentially multiple concurrent sound sources, so

$$H_{S}(\Theta_{s},\theta,\phi) = \sum_{s=1}^{S} A_{s}g(\Theta_{s},\theta,\phi)$$
(6)

represents a sum of contributions to the model over some number S of sources, where A_s is the amplitude of each contribution and g provides its shape and direction. Specifically,

$$g(\Theta_s, \theta, \phi) = 2\pi \sum_{n=1}^N \sum_{m=-n}^n Y_n^{m*}(\Theta_s) Y_n^m,$$
(7)

where $\Theta_S = \{\theta_1, ..., \theta_S; \phi_1, ..., \phi_S\}$ denotes the directions of the modeled sources. This equation is, itself, equivalent to that of a beam pattern listening in direction Θ_s . [8]

3. MODEL-BASED BAYESIAN ANALYSIS

This work uses Bayes' Theorem in two distinct but interrelated tasks: model selection and parameter estimation. Model selection in this case consists of selecting the right number of sound sources to include in the model, while parameter estimation involves choosing the angles and amplitudes for each source. If $\mathbf{H}_i = [H_i(\Theta_i, \theta, \phi)]$ is allowed to represent a particular model (see equation 6) and $\mathbf{D} = [D(\theta, \phi)]$ represents some measured data set (see equation 3), then Bayes' Theorem can be stated as

$$p(\mathbf{H}_i|\mathbf{D}) = \frac{p(\mathbf{D}|\mathbf{H}_i)p(\mathbf{H}_i)}{p(\mathbf{D})},$$
(8)

where $p(\cdot)$ represents probability and p(a|b) represents the probability of *a* given *b*. This equation describes model selection, the first tier of Bayesian inference employed. $p(\mathbf{H}_i|\mathbf{D})$, the probability of the model being true given some data, is expressed as a function of $p(\mathbf{H}_i)$, the prior probability of the model based on previous knowledge (assigned uniformly according to the principle of maximum entropy), $p(\mathbf{D})$, the probability of observing the data which simply acts as a normalizing constant, and $p(\mathbf{D}|\mathbf{H}_i)$, the evidence term, representing the likelihood of the data occurring should the model be true. [3]

Likewise, the second tier of inference can be expressed as

$$p(\boldsymbol{\Theta}|\mathbf{D},\mathbf{H}) = \frac{p(\mathbf{D}|\boldsymbol{\Theta},\mathbf{H})p(\boldsymbol{\Theta}|\mathbf{H})}{p(\mathbf{D}|\mathbf{H})},$$
(9)

where the subscripts have been dropped for simplicity. Here, the probability of some parameter set $\Theta = \{\theta; \phi; \mathbf{A}\}$ is to be determined for some already chosen model **H**. Since the probabilities are normalized and $p(\mathbf{D}|\mathbf{H})$ is independent of Θ , equation 9 can be integrated over the entire parameter space to yield

$$p(\mathbf{D}|\mathbf{H}) = \int_{\Theta} p(\mathbf{D}|\Theta, \mathbf{H}) p(\Theta|\mathbf{H}).$$
(10)

Next, $p(\mathbf{D}|\mathbf{\Theta}, \mathbf{H})$ can be found as a function of the squared error between D and H, where both are two-dimensional arrays over θ and ϕ determined above. [6] With $p(\mathbf{D}|\mathbf{\Theta}, \mathbf{H})$ in hand, equation 10 can be used to determine $p(\mathbf{D}|\mathbf{H})$, since $p(\mathbf{\Theta}|\mathbf{H})$ is prior knowledge. Thus equations 8 and 9 can then be evaluated to find the posterior probabilities for the models and their parameters, respectively. By this method, both inference steps can be completed using only the comparison of modeled data to experimental data. This comparison is carried out in a random nested sampling scheme to be presented in more detail in an upcoming publication. [6]

Once the posterior probabilities, $p(\mathbf{H}|\mathbf{D})$, for each model have been determined, they can be compared using a Bayes factor relating the probability of two adjacent models to one another. This can be expressed logarithmically in decibans by

$$L_{ij} = 10 \log_{10} \left(\frac{p(\mathbf{H}_i | \mathbf{D})}{p(\mathbf{H}_j | \mathbf{D})} \right).$$
(11)

This factor is used to compare the probability of one model (number of sources) to the previous one as the number of sources is incremented. The model that yields the greatest Bayes factor is considered to be the correct one. This selection criterion serves as an implementation of Occam's razor, favoring the simplest model that can adequately fit the data without overfitting.

4. EXPERIMENTAL RESULTS

Some preliminary results are included below for data collected from three sound sources (S = 3). Measurements of three impulse responses were taken with the

loudspeaker source located between 1 and 1.5 meters from the spherical array receiver. The receiver was a rigid sphere, with sixteen microphones sampling its surface in a nearly-uniform arrangement. The three impulse responses were recorded sequentially, then added, after windowing out room reflections, in order to represent multiple simultaneous sources. The source directions are $(5^\circ, 60^\circ)$, $(135^\circ, 140^\circ)$, and $(270^\circ, 90^\circ)$. Finally, white noise was convolved with the combined impulse responses to create the signal to be analyzed.

Equation 3 is then applied to these pressure data, with the look direction swept over all angles. The results of this process are shown in figure 1 (a). The Bayesian model selection process described above is then executed, ultimately selecting the model shown in 1 (b). The directions of arrival results from this model selection are displayed in table 1. The model selection process was repeated fifteen times in order to evaluate its consistency.



Figure 1: Directional mapping of sound energy over the sphere for three sound sources located at $(5^{\circ}, 60^{\circ})$, $(135^{\circ}, 140^{\circ})$, and $(270^{\circ}, 90^{\circ})$. (a) Representation of experimental data generated by applying the beamforming equation 3 over all angles to the measured pressure data. (b) Analytic model determined by the Bayesian model selection process to fit the data.

Figure 2 displays the mean evidence for each number of sources, up to five. Note that while the three source model did not have the highest evidence in absolute terms, it did have the highest Bayes factor. As such, it was selected as the correct model because more complex models did not offer the same degree of improvement in evidence, indicating that the further complexity was unwarranted.

5. DISCUSSION AND CONCLUSION

By simple inspection, it is not possible to locate sound sources in figure 1. For a lower number of sources, it may be possible to locate sources by correlating them with maxima

Table 1: Results of the DoA determination algorithm. True position is the position of the speaker when the signal was recorded, estimated position is the position determined by the Bayesian model selection process, estimate deviation is the total variance between the fifteen runs of the algorithm, and the error refers to the difference in azimuth and elevation between the estimated position and the true position.

Source Number:	1	2	3
True Position	(5°, 60°)	(135°, 140°)	(270°, 90°)
Estimated Position	(8.7°, 72.4°)	(125.8°, 148.6°)	(254.1°, 71.5°)
Estimate Deviation	$(\pm 8.6^{\circ}, \pm 8.2^{\circ})$	$(\pm 17.5^{\circ}, \pm 1.4^{\circ})$	$(\pm 7.7^{\circ}, \pm 7.4^{\circ})$
Error	(3.7°, 12.4°)	(9.2°, 8.6°)	(15.9°, 18.5°)



Figure 2: (a) Comparison of the mean Bayesian evidence for each number of sources, one to five. The error bars represent the total variation in evidence observed over fifteen repetitions of the algorithm. (b) Bayes factor, comparing the current model to the previous one.

in sound energy, but as the number of sources grows, this becomes impractical. However, the Bayesian model selection process employed in this work was able to make some sense of the data and make reasonable estimates of the source locations. This suggests that the two-tiered system of Bayesian analysis is capable of accomplishing its tasks of determining the number and locations of sound sources.

However, a certain degree of ambiguity in the results cannot be denied, and this even with the elimination of room reverberation from the data. The moderate inaccuracy and imprecision of the results seen in table 1 are likely the result of the limited resolution of the sixteen-channel array employed in this work. Due to this resolution, spherical harmonics of only up to second order could be used to describe the sound field. With this in mind, further investigation is planned using a 32-channel spherical array. This should enable

higher order beamforming, which in turn should allow for greater angular precision. The influence of this precision on the behaviour of the model selection scheme is to be determined. Additionally, the performance of the algorithm with reverberant energy present in the recorded sound field remains to be characterized. It is hoped that a higher order microphone array will be able to achieve more precise models of higher numbers of sound sources and that the probabilistic model selection process will be sufficiently robust to be able to count and localize these sources in a more complex acoustical environment. However, these remain subjects for future research.

6. **REFERENCES**

- [1] J. Meyer and G. Elko. A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield. In 2002 IEEE Int. Conf. on Acoust., Speech and Signal Process., volume II, pages 1781–1784.
- [2] B. Rafaely. *Fundamentals of Spherical Array Processing*. Springer-Verlag GmbH Berlin Heidelberg, Berlin, 2015.
- [3] C. Fackler N. Xiang. Objective bayesian analysis in acoustics. *Acoustics Today*, 11(2):55–61, 2015.
- [4] D. Bush and N. Xiang. A model-based bayesian framework for sound source enumeration and direction of arrival estimation using a coprime microphone array. *J. Acoust. Soc. Am.*, 143(6):3934–3945, June 2018.
- [5] J. Escolano, N. Xiang, J. Perez-Lorenzo, M. Cobos, and J. Lopez. A bayesian direction-of-arrival model for an undetermined number of sources using a two-microphone array. J. Acoust. Soc. Am., 135(2):742–753, February 2014.
- [6] C. Landschoot and N. Xiang. Model-based bayesian direction of arrival analysis for sound sources using a spherical microphone array. *Submitted to J. Acoust. Soc. Am.*, 2019.
- [7] E. Williams. *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography.* Academic Press, London, 1999.
- [8] Z. Li and R Duraiswami. Hemispherical microphone arrays for sound capture and beamforming. In 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pages 106–109, 2005.