PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Bayesian inference in direction of arrival analysis using spherical microphone arrays

Ning XIANG [1]; Stephen WEIKEL[2]; and Christopher LANDSCHOOT [2]

[1,2] Graduate Program in Architectural Acoustics, Rensselaer Polytechnic Institute, Troy, New York, 12180 USA

ABSTRACT

One challenging problem in acoustical applications using microphone arrays is determining the directions of arrivals (DoAs) of multiple sound events. This work applies a unified Bayesian framework to address this problem in situations involving potentially multiple simultaneous sound sources using spherical microphone arrays. It presents a two-level inferential problem of sound source enumeration and direction of arrival estimation. To this end, analytical models based on spherical harmonics are used to predict experimental data collected by spherical microphone arrays. The prediction is evaluated against the measured data in order to select the simplest such model that can adequately match the experimental data, thereby estimating first the number of sources, then their DoA information. This paper presents the analytic models, the two levels of Bayesian formulation, and analysis results to demonstrate the potential usefulness of this model-based Bayesian analysis for complex sound environments with potentially multiple concurrent sources.

Keywords: sound source localization, direction of arrivals, Bayesian analysis, spherical harmonics.

1. INTRODUCTION

This paper addresses a twofold challenge of determining both the number and location(s) of the detected sound sources. Collection of sound signals relies on a spherical microphone array. The data are used to extract directional information via spherical harmonic beamforming. Bayesian analysis is used to enumerate first the number of concurrent sources, followed by their strengths and directions of arrivals (DoAs). The spherical beamforming technique applies specific weighting factors to the detected sound pressure around the surface of the spherical microphone array [1,2]. The process creates a map of arrival directions of incident sound energy. This work investigates the capabilities of two-level Bayesian inference to carry out these steps. [3] Similar work using Bayesian inference has been reported in the literature to determine source number and position. However, the data are collected using different microphone arrays [4,5]. To the best of the authors' knowledge, this current work is the first time that Bayesian analysis is applied to analyze spherical microphone data. This paper will present a brief description of the beamforming techniques and model formulation applied in the two-levels of Bayesian inference followed by discussing some experimental results.

[1] xiangn@rpi.edu

A more thorough discussion of the work is documented in an upcoming journal publication [6].

2. BEAMFORMING DATA and MODEL

Sound energy impinging upon a spherical microphone array can be processed using the spherical Fourier transform [2] of the microphone signals, $p(k, r, \theta_i, \phi_i)$

$$y(\theta_l, \phi_l) = \frac{1}{M} \sum_{n=1}^{N} \sum_{m=-n}^{n} \left\{ w_{nm}^{*}(k, r, \theta_l, \phi_l) \sum_{i=1}^{M} p(k, r, \theta_i, \phi_i) \left[Y_n^m(\theta_i, \phi_i) \right]^{*} \right\}, \quad (1)$$

where M is the number of microphones, θ_i and ϕ_i represent the angular position of the ith microphone, k is the wave number, r is the radius of the sphere, and * denotes the complex conjugate. And $Y_n^m(\theta_i, \phi_i)$ is the spherical harmonics of order n and degree m [2]. When specifying some looking direction θ_l, ϕ_l, spatially filtered pressure signals can be expressed by applying beamforming weights $w_{nm}^{*}(k, r, \theta_l, \phi_l)$ to the above transformed pressure signal. The maximum order of spherical harmonics, N, is dictated by the number of microphones built in the spherical array, and

$$w_{nm}^{*} = \frac{Y_n^m(\theta_l, \phi_l)}{\mathrm{j}^n b_n(kr)}, \quad (2)$$

with $\mathrm{j}^n = (-1)^{n/2}$ and $b_n(kr)$ representing spherical modal amplitude given by

$$b_n(kr) = j_n(kr) - \frac{j_n'(kr)}{h_n^{(2)'}(kr)} h_n^{(2)}(kr), \quad (3)$$

where j_n represents the spherical Bessel function of the first kind and $h_n^{(2)}$ represents the Hankel function of the second kind. The prime denotes the derivative with respect to the argument [2]. r is the radius of the rigid sphere. Eventually, normalized sound energy over a predefined angular range can be determined

$$D(\theta_l, \phi_l) = \frac{y^2(\theta_l, \phi_l)}{\max[y^2(\theta_l, \phi_l)]}. \quad (4)$$

The beamforming data can be predicted solely using the spherical harmonics

$$H_S(\Theta_S, \theta, \phi) = \sum_{s=1}^{S} A_s \frac{g^2(\Theta_s, \theta, \phi)}{\max[g^2(\Theta_s, \theta, \phi)]}, \quad (5)$$

where, the model, H_S, describes a sum of potentially multiple (S) simultaneous sound sources, $\Theta_S = \{\theta_1, \phi_1, \cdots \theta_S, \phi_S\}$. A_s is amplitude of each sound source, and

$$g(\Theta_s, \theta, \phi) = 2\pi \sum_{n=1}^{N} \sum_{m=-n}^{n} Y_n^m(\theta, \phi) \left[Y_n^m(\Theta_s) \right]^*, \quad (6)$$

where $\Theta_s = \{\theta_s, \phi_s\}$, denotes the direction of a modeled sound source. Equation (6) describes a beam pattern focused in direction Θ_s [6].

3. BAYESIAN ANALYSIS

This work applies two-levels of Bayesian inference to solve the direction of arrival (DoA) problems. The higher level of inference represents model selection, while the lower one parameter estimation. The model selection in this work evaluates models describing the number of simultaneous sound sources, while parameter estimation estimates the angles and amplitudes for each source. Detailed description can be found in the upcoming journal publication by Landschoot and Xiang [6].

At this stage a unified Bayesian framework can be formulated as

$$p(\Theta_s \mid D, H) \quad \cdot \quad Z \quad = \quad L(\Theta_s) \quad \cdot \quad \pi(\Theta_s). \quad (7)$$

Posterior evidence = likelihood prior

The likelihood represents the probability of mismatch between the data in eq.(4) and the modelled prediction in eq.(5). This work applies the principle of maximum entropy (MaxEnt) to assign this probability, which leads to Student-t distribution as

$$L(\Theta_s) \propto \Gamma(\frac{Q}{2}) \frac{(2\pi E)^{-Q/2}}{2}, \quad (8)$$

where Q is the total number of data points and $\Gamma(\cdot)$ is the Gamma function. The MaxEnt also assigns the prior probability, $\pi(\Theta_s)$, of angular variables to be uniform. This work has to first estimate the evidence, Z, in order to select a correct model. The following step is carried out by estimating the Bayes factor in logarithmic form as

$$B_{ij} = 10 \log_{10} Z_i - 10 \log_{10} Z_j. \quad (9)$$

where Z_i is the evidence of the beamforming model containing i number of sound sources. The evidence Z_i is determined through exploration of the likelihood over the prior as

$$Z_i = \int_{\Theta_i} L(\Theta_i) \pi(\Theta_i) \, d\Theta_i. \quad (10)$$

Once the evidence of Z_i and the evidence Z_j are estimated, the Bayes factor in eq.(9) can be determined, expressing how much the data prefer model H_i over model H_j. Using the Bayes factor, one of the models will be selected, say, Z_S. The lower level of inference can be pursued, eq.(7) yields the posterior probability,

$$p(\Theta_S \mid D, H_S) = \frac{L(\Theta_S)\,\pi(\Theta_S)}{Z_S}. \tag{11}$$

During the exploration of the likelihood over the prior probability throughout the entire parameter space, the likelihood is sufficiently explored, so that the posterior probability is already approximately available. This work uses nested sampling for both levels of inference.

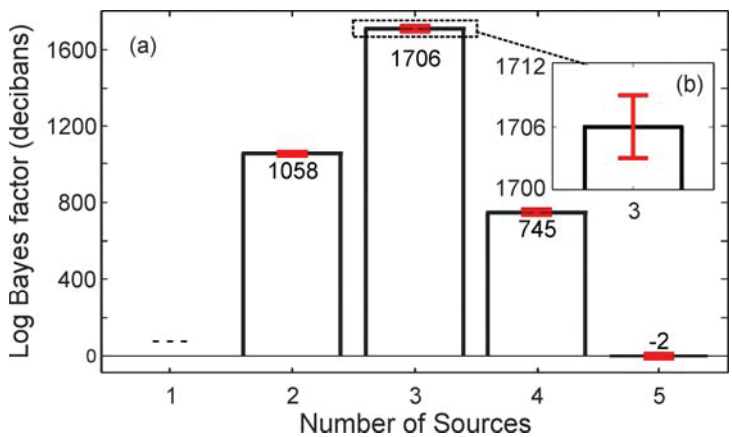

Figure 1 -- Bayes factors evaluated for the number of simultaneous sound sources. (a) Bayes factors from 1 to 5 sources. 3 over 2 sources indicates unambiguous preference to the three source model over that of 2 sources, much higher than the Bayesian factor of 2 source model over 1 source and higher than that of 4 over 3 sources. (b) Magnified view of variation for Bayes factor of three over two sources.

4. EXPERIMENTAL RESULTS

This section discusses some preliminary results. Details can be found in the journal publication by Landschoot and Xiang [6]. In order to prove the concept stated previously, experimental data collected from three sound sources (S = 3) are used for validating the theory. Experimental measurements of three room impulse responses were taken with the loudspeaker source located between 1 and 1.5 meters from the spherical microphone array. Sixteen microphones sampling a rigid sphere with its surface in a nearly-uniform arrangement. The source directions are (5^{o}; 60^{o}), (135^{o}; 140^{o}), and (270^{o}; 90^{o}). When white noise segments are convolved with the direct-sound portions of three impulse responses and then combined linearly, the sound energy in eq.(4) is created. The Bayes factors as illustrated in Fig.1 indicate unambiguous preference to the three source model over that of two sources, much higher than the Bayesian factor of four over three sources, and much higher than that of two source model over one source and so on. Figure 2 compares the sound energy map between experimental data shown in eq.(4) and the Bayesian model prediction via the final parameter estimation in eq.(5). Figure 2 indicates simply correlating sound energy peaks will fail to estimate directions of arrivals for this case of three simultaneous sound sources.

5. SUMMARY

The present work applies a Bayesian method to beamformed models, evaluating them against experimental data that is recorded by a spherical microphone array in order to estimate the directions of arrivals (DoAs) of simultaneous sound sources. Through the two-level Bayesian inference approach to this problem involving first estimating the number of sound sources and second estimating their DoAs, both of these pieces of information can be reliably estimated. This Bayesian inference approach provides an improvement in the detection of sound sources over previous methods, such as those that directly correlate the peak energies to the DoAs. This work demonstrates the feasibility of Bayesian model selection as a means to determine the DoAs of sound sources. Nevertheless, there are still challenges that can be addressed in future research. For future room acoustical applications, e.g., this method of DoA analysis could potentially be extended to determine the locations of reflections within an enclosed space. Since room impulse responses can be acquired with a spherical array in an enclosed space, the data gathered in such a measurement could be used to identify the origins of troublesome reflections.

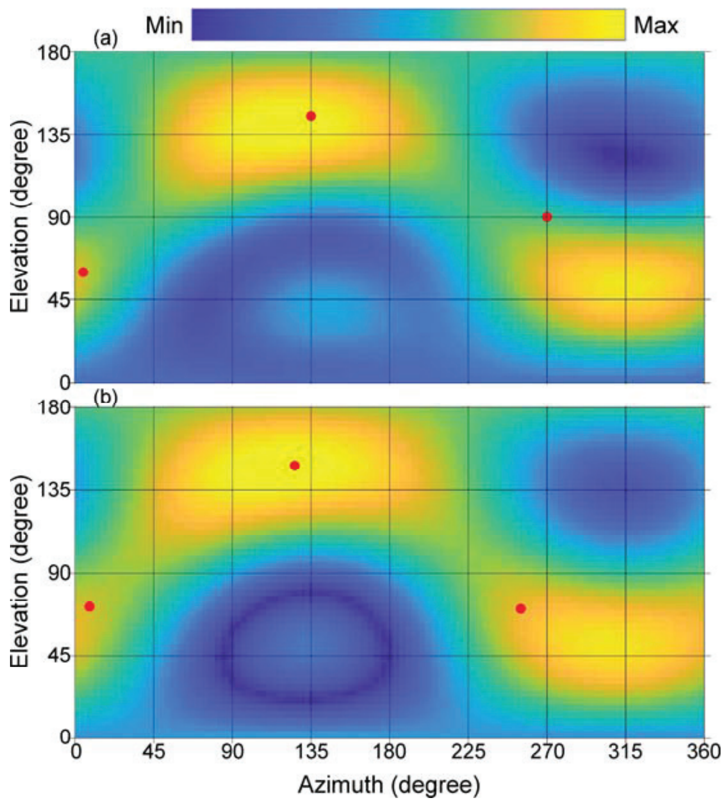

Figure 2 -- Comparison between the experimental data and the model prediction. After model selection the model of three sound sources is used for the parameter estimation of directions of arrives from three sources. Three solid dots indicate the source positions. (a) Experimental data. (b) Model prediction.

REFERENCES

1. Meyer J. and Elko, G. A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield, *2002 IEEE Int. Conf. on Acoust., Speech & Signal Process.*, **Vol. II**, 1781–1784 (2002).
2. Rafaely, B. *Fundamentals of Spherical Array Processing*, Springer-Verlag GmbH Berlin Heidelberg, Berlin, 2015.
3. Xiang, N. and Fackler C. J. Objective Bayesian analysis in acoustics, *Acoustics Today*, **11**, 55–61 (2015).
4. Bush, D. and Xiang, N. A model-based Bayesian framework for sound source enumeration and direction of arrival estimation using a coprime microphone array, *J. Acoust. Soc. Am.*, **143**, 3934–3945 (2018).
5. Escolano, J. Xiang, N. Perez-Lorenzo, J. Cobos, M. and Lopez, J. A Bayesian direction-of-arrival model for an undetermined number of sources using a two microphone array, *J. Acoust. Soc. Am.*, **135**, 742–753 (2014).
6. Landschoot, C. and Xiang, N. Model-based Bayesian direction of arrival analysis for sound sources using a spherical microphone array, *J. Acoust. Soc. Am.*, **145**, (to print) 2019.