The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Information-Theoretic Detection of Broadband Sources in a Coherent Beamspace MUSIC Scheme

Patrick Danès and Julien Bonnal

Abstract— **This paper deals with the detection of the number of broadband sources in an acoustic environment from a small-size, embeddable microphone array. The coupling of information-theoretic statistical identification methods with MUSIC schemes, which has long been acknowledged in the array processing community, is first reviewed. From these considerations, a source number detector based either on the Akaike Information Criterion** (AIC) **or on the Minimum Description Length** (MDL) **criterion is derived within an original coherent broadband beamspace MUSIC method. This constitutes a theoretically sound solution, well-suited to robotics as it requires no subjective threshold setting and has a limited computational cost. Experimental results validate the approach. All the necessary theoretical background is provided, so that the paper is self-contained.**

I. INTRODUCTION

The data of the number of sources is often a prerequisite to the processing of auditory signals. Many localization or separation algorithms rely on this prior knowledge, and their robustness to wrong assumptions may be very poor. Particularly in robot audition, source detection turns out to be a key issue. With the tightly connected localization stage, it constitutes a fundamental precondition to many higher-level functions [1]. In addition, the number of active sources can change within the uncontrolled and open-ended environments of robotics, which makes the problem even more challenging.

So, source detection has been investigated since long ago, *e.g.* in [2] through a probabilistic processing downstream a beamforming based localization. Ref. [3] estimates the time delays of arrival by an eigenstructure-based generalized cross-correlation method, then detects the source number by applying an adaptive K-means algorithm on the deduced least squares approximations of the source directions and velocities. MUSIC [7], which constitutes a framework for many perspectives [4], has been considered for detection in [5]. Therein, narrowband MUSIC schemes are applied on a frequency decomposition of the sensed signals, assuming a common signal space dimension. Then, a higher maximum number of sources is allowed during the isolation of the peaks of the—average—broadband pseudo-spectrum. Nevertheless, an empirical tuning of parameters is required.

Despite many contributions, a theoretically sound and computationally efficient solution to source number detection still needs to be devised. This paper aims at providing some elements towards such a design, first by bringing to the fore theoretical concepts which have long been acknowledged in the array processing community, then by instantiating them into an original approach well-suited to robotics. It is organized as follows. Section II states the problem, recalls elements of the MUSIC localization algorithm, and discusses basic detection schemes. Sections III and IV then thoroughly explain how the introduction of information-theoretic criteria enable the detection of narrowband and broadband sources, respectively. An information-theoretic detector within a coherent beamspace MUSIC follows, which fits the requirements of robotics. Experiments constitute Section V. A conclusion ends the paper.

This work was supported by the French ANR AMORCES and the EU FP6-STREP CommRob projects.

P. Danès and J. Bonnal are with CNRS; LAAS; 7 avenue du colonel Roche, F-31077 Toulouse, France; and with Université de Toulouse; UPS, INSA, INP, ISAE; LAAS; F-31077 Toulouse, France {patrick.danes,julien.bonnal}@laas.fr

II. FUNDAMENTAL ISSUES

Notations are standard. Normal (resp. bold) lowercase/uppercase letters depict scalar (resp. vector) signals in the time/frequency domains. Underlined bold letters relate to matrices. The $N \times N$ identity and $N \times M$ zero matrices are denoted $\underline{\mathbb{I}}_N$ and $\underline{\mathbb{O}}_{N,M}$, and their subscripts may be omitted. The transpose and Hermitian transpose operators are respectively termed $(.)^T$ and $(.)^H$. $\mathbb{E}[.]$ stands for the expectation operator, $\triangleq$ means "equal to, by definition", and $\mathcal{CN}(\mathbf{z};\bar{\mathbf{z}},\underline{\mathbf{C}}_{\mathbf{z}})$ denotes the complex Gaussian distribution on $\mathbf{z}$ with mean $\bar{\mathbf{z}}$ and covariance $\underline{\mathbf{C}}_{\mathbf{z}}$.

A. Problem Statement

An array of N omnidirectional sensors samples a wavefield propagating at velocity c from $D < N$ pointwise broadband sources. These lie in the farfield, so that their locations are depicted by their azimuths and elevations relative to a frame $\mathscr{F}$ linked to the array. At $\mathscr{F}$'s origin, the wavefield is thus the sum of D contributions $s_1(t),\ldots,s_D(t)$, one per source. The complex envelopes of $\mathbf{s}(t) \triangleq (s_1(t),\ldots,s_D(t))^T$ constitute $\mathbf{S}(k) = (S_1(k),\ldots,S_D(k))^T \in \mathbb{C}^D$, with $k = \frac{2\pi f}{c}$ and f the spatial and temporal frequencies. Let $\mathbf{V}(\mathbf{r}_d,k) \in \mathbb{C}^N$ stand for the steering vector of the d^{th} source at azimuth and elevation depicted by $\mathbf{r}_d$. Then, the complex envelopes $\mathbf{N}(k) = (N_1(k),\ldots,N_N(k))^T$ and $\mathbf{X}(k) = (X_1(k),\ldots,X_N(k))^T$ of the noise $\mathbf{n}(t) \triangleq (n_1(t),\ldots,n_N(t))^T$ and signal $\mathbf{x}(t) \triangleq (x_1(t),\ldots,x_N(t))^T$ at the N sensors write as

$$\mathbf{X}(k) = \underline{\mathbf{V}}(\mathbf{r}_1,\ldots,\mathbf{r}_D,k)\mathbf{S}(k) + \mathbf{N}(k), \tag{1}$$

with $\underline{\mathbf{V}}(\mathbf{r}_1,\ldots,\mathbf{r}_D,k) = \left(\mathbf{V}(\mathbf{r}_1,k) \mid \ldots \mid \mathbf{V}(\mathbf{r}_D,k)\right)$ [6]. The function $\mathbf{V}(.,.)$ is supposed known analytically or by array calibration. The processes $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are assumed mutually independent, zero-mean, stationary and ergodic on the time window of interest. $\underline{\mathbf{C}}_{\mathbf{s}}(k) \triangleq \mathbb{E}[\mathbf{S}(k)\mathbf{S}^H(k)]$ is unknown, and possibly

singular at every frequency k as in the case of severe mutipath propagation. $\mathbf{n}(t)$ can have any frequency contents and spatial distribution, and $\underline{\mathbf{C}}_{\mathbf{n}}(k) \triangleq \mathbb{E}[\mathbf{N}(k)\mathbf{N}^H(k)] = \sigma_{\mathbf{n}}^2 \underline{\mathbf{C}}_{\mathbf{n}}(k)$ is known up to the constant $\sigma_{\mathbf{n}}^2$. $\underline{\mathbf{V}}(\mathbf{r}_1,\ldots,\mathbf{r}_D,k)$ must be full-rank at $\mathbf{r}_1,\ldots,\mathbf{r}_D$, which is generally true if $\mathbf{r}_1 \neq \mathbf{r}_2 \neq \cdots \neq \mathbf{r}_D$.

The problem is then to *detect*—i.e. *estimate the number D of—the active sources* from the data of $\mathbf{x}(t)$. To simplify, the array is supposed linear with uniform interspace, and $\mathscr{F}$'s origin is set to its midpoint O. By the rotational symmetry around the array axis, each dependency on the angular coordinate vector $\mathbf{r}_d$ of a d^{th} source reduces to a dependency on its azimuth—or bearing—θ_d. The n^{th} entry of the steering vector at any dummy azimuth θ measured with respect to endfire thus reads as

$$V_n(\theta,k) = e^{jkz_n\cos\theta},\ n = 1,\ldots,N, \tag{2}$$

with z_n the distance of the n^{th} microphone to O.

B. The Elementspace MUSIC Method to the Detection of Narrowband Farfield Sources

The celebrated MUltiple SIgnal Classification (MUSIC) method not only enables the high-resolution localization of sources, but also offers a sound framework to their detection. A simplified "narrowband" problem is stated first. The basic "elementspace" MUSIC algorithm—*i.e.* in the space of the array elements—is outlined next [7].

1) A Simplified Problem: To begin with, the D sources are supposed narrowband with common center frequency k_0, and not coherent with each other. So, their covariance matrix $\underline{\mathbf{C}}_{\mathbf{s}} \triangleq \mathbb{E}[\mathbf{s}(t)\mathbf{s}^H(t)] = \underline{\mathbf{C}}_{\mathbf{s}}(k_0)$ is full rank. The noise $\mathbf{n}(t)$, independent of $\mathbf{s}(t)$, is spatially white, so that $\mathbb{E}[\mathbf{s}(t)\mathbf{n}^H(t)] = \underline{\mathbb{O}}$ and $\mathbb{E}[\mathbf{n}(t)\mathbf{n}^H(t)] = \sigma_{\mathbf{n}}^2\underline{\mathbb{I}}_N$, with unknown $\sigma_{\mathbf{n}}^2$.

2) Elementspace Narrowband MUSIC to Source Detection and Localization: In this simplified problem, the covariance matrix $\underline{\mathbf{C}}_{\mathbf{x}} \triangleq \mathbb{E}[\mathbf{x}(t)\mathbf{x}^H(t)] = \underline{\mathbf{C}}_{\mathbf{x}}(k_0)$ has the form

$$\underline{\mathbf{C}}_{\mathbf{x}} = \underline{\mathbf{V}}(\theta_1,\ldots,\theta_D,k_0)\underline{\mathbf{C}}_{\mathbf{s}}\underline{\mathbf{V}}^H(\theta_1,\ldots,\theta_D,k_0) + \sigma_{\mathbf{n}}^2\underline{\mathbb{I}}_N. \tag{3}$$

Its eigendecomposition then writes as

$$\underline{\mathbf{C}}_{\mathbf{x}} = \sum_{i=1}^{N} \lambda_i \mathbf{U}_i\mathbf{U}_i^H, \tag{4}$$

where its eigenvalues satisfy $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_D > \lambda_{D+1} = \ldots = \lambda_N = \sigma_{\mathbf{n}}^2$ and can be associated with orthogonal right eigenvectors $\mathbf{U}_1,\ldots,\mathbf{U}_N$. In addition, the columns of $\underline{\mathbf{U}}_{\mathscr{S}} = (\mathbf{U}_1 \mid \ldots \mid \mathbf{U}_D) \in \mathbb{C}^{N\times D}$ span the so-called "signal space" $\mathscr{S} \subset \mathbb{C}^N$ generated by the steering vectors $\mathbf{V}(\theta_1,k_0),\ldots,\mathbf{V}(\theta_D,k_0)$ at the source azimuths. The orthogonal complement $\mathscr{N}$ of $\mathscr{S}$, also termed "noise space", is then the range of the matrix $\underline{\mathbf{U}}_{\mathscr{N}} = (\mathbf{U}_{D+1} \mid \ldots \mid \mathbf{U}_N) \in \mathbb{C}^{N\times(N-D)}$ made of the remaining eigenvectors associated to $\lambda_{D+1} = \ldots = \lambda_N = \sigma_{\mathbf{n}}^2$.

Consequently, even if $\sigma_{\mathbf{n}}^2$ is unknown, the number D of active sources is computed as N minus the multiplicity of the smallest eigenvalue of the array spatial covariance matrix $\underline{\mathbf{C}}_{\mathbf{x}}$. Then, the so-called—theoretical—"pseudo-spectrum"

$$h(\theta,k_0) \triangleq \frac{1}{\mathbf{V}^H(\theta,k_0)\underline{\Pi}_{\mathscr{N}}\mathbf{V}(\theta,k_0)}, \tag{5}$$

with $\underline{\Pi}_{\mathscr{N}} \triangleq \sum_{i=D+1}^{N} \underline{\mathbf{U}}_i\underline{\mathbf{U}}_i^H$ the "projector on the noise space", shows infinite peaks at the source bearings, *i.e.* iff $\theta \in \{\theta_1,\ldots,\theta_D\}$.

3) The Colored Noise Case: The method straightly applies to the case when the noise, though still independent of the sources, is spatially correlated. If $\mathbb{E}[\mathbf{n}(t)\mathbf{n}^H(t)] = \sigma_{\mathbf{n}}^2\underline{\mathbf{C}}_{\mathbf{n}}$ is known up to the constant $\sigma_{\mathbf{n}}^2$, then the detection and localization are unchanged provided $\{\lambda_i,\mathbf{U}_i\}_{i=1,\ldots,N}$ term the generalized eigenvalues and eigenvectors of the matrix pencil $(\underline{\mathbf{C}}_{\mathbf{x}},\underline{\mathbf{C}}_{\mathbf{n}})$. The relationship $\mathbf{V}^H(\theta,k_0)\underline{\mathbf{U}}_{\mathscr{N}} = \mathbf{0}^T$ underlying the definition (5) still holds, but the generalized eigenvectors satisfy $(\underline{\mathbf{U}}_{\mathscr{S}} \mid \underline{\mathbf{U}}_{\mathscr{N}})^H\underline{\mathbf{C}}_{\mathbf{n}}(\underline{\mathbf{U}}_{\mathscr{S}} \mid \underline{\mathbf{U}}_{\mathscr{N}}) = \underline{\mathbb{I}}_N$.

4) MUSIC in Practice: In practice, the genuine covariance matrix $\underline{\mathbf{C}}_{\mathbf{x}}$ is not available in closed form, and a finite-sample estimate $\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}$ is computed instead. Therefore, an approximated projector $\tilde{\underline{\Pi}}_{\mathscr{N}}$ is entailed in (5) in place of $\underline{\Pi}_{\mathscr{N}}$, leading to a practical pseudo-spectrum $\tilde{h}(\theta,k_0)$ showing sharp finite peaks at the source locations. If the estimate $\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}$ is asymptotically unbiased, then so are the MUSIC bearing estimates. In the general case when the number of active sources is not known beforehand, the mismatch between $\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}$ and $\underline{\mathbf{C}}_{\mathbf{x}}$ almost surely implies that the $N-D$ smallest eigenvalues are distinct. This in turn may hinder the separation of the signal and noise spaces, and may induce a significant performance drop, see *e.g.* [8].

C. Two Basic but Unworkable Source Detection Schemes

Some MUSIC-based detectors are hereafter discussed.

1) Ruling out the Most Elementary Detection Scheme: One can wonder about first defining $N-1$ separate MUSIC pseudo-spectra $\tilde{h}(\theta,k_0|d)$, each one corresponding to an hypothesized number of sources $d \in \{0,\ldots,N-1\}$, prior to detecting $\hat{d} = \arg\max_d\big(\max_\theta \tilde{h}(\theta,k_0|d)\big)$. Such an approach has no sound theoretical basis and can be straightly ruled out. Indeed, $\tilde{h}(\theta,k_0|d)$ is the inverse of the quadratic Euclidean distance of $\mathbf{V}(\theta,k_0)$ to the signal space defined from the sample estimate of the covariance matrix under the assumption of d sources. So, $\max_\theta \tilde{h}(\theta,k_0|d)$ is always maximum for the highest-dimension signal space, *i.e.* for $\hat{d} = N-1$.

2) An Hypothesis Testing Based Approach: The oldest rigorous source detection method relies on the sphericity test [9], which checks for the sphericity of the iso-density contours of a M-dimensional Gaussian random vector $\mathbf{r}$. In other words, given a sample approximation $\tilde{\underline{\mathbf{R}}}$ of $\mathbf{r}$'s genuine covariance matrix $\underline{\mathbf{R}}$, this test deduces if the reference hypothesis H_0:*all the eigenvalues of* $\underline{\mathbf{R}}$ *are equal* better explains the distribution of $\tilde{\underline{\mathbf{R}}}$'eigenvalues than the alternative hypothesis H_1:*the extremum eigenvalues of* $\underline{\mathbf{R}}$ *are distinct.* Its implementation takes the form of a Generalized Likelihood Ratio Test (GLRT), whose threshold should be deduced from a probability of false alarm selected beforehand.

Its application to source detection consists in a nested series of such binary hypotheses, each of which checks for equality of an increasing $\tilde{d}$-element subset of the smallest eigenvalues of $\underline{\mathbf{C}}_{\mathbf{x}}$ from its sample approximation $\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}$. The detected number of sources is then the maximum value of $\tilde{d}$ below which this series of tests passes. Some fundamental

drawbacks however limit the pertinence of this approach. First, given a desired false alarm probability, it is impossible to determine a threshold for the whole detection problem. At best, a threshold can be rigorously defined for each nested GLRT, despite the knowledge of the statistics of the sample covariance eigenvalues under the H_0 hypothesis—which is necessary to bridge the gap with the desired probability of false alarm—is hard to get in closed form. Secondly, the difficulty to characterize the statistical dependence of the individual tests prevents any deduction of the false alarm probability of the global detection problem.

III. INFORMATION-THEORETIC DETECTION OF NARROWBAND FARFIELD SOURCES

In light of the above negative comments on source detection based on statistical hypotheses tests, the exploitation of information-theoretic criteria is henceforth presented, which is the main topic of the paper. Statistical identification is first reviewed, with a special focus on the case of competing models. A procedure taking the form of a criterion minimization is reviewed which, contrarily to conventional hypotheses testing, does not require any subjective threshold setting. The way how it can be declined into constructive algorithms to the detection of narrowband sources then follows.

A. A Bird's View at Statistical Identification

1) Basics: Consider a vector random variable $\mathbf{y}$ with probability density function $g(\mathbf{y})$, and $f(\mathbf{y}|\rho)$ a density function with vector parameter ρ used as a model for $g(\mathbf{y})$. The Kullback-Leibler divergence $\mathscr{K}(g;f(.|\rho)) \triangleq \int g(\mathbf{y}) \ln \frac{g(\mathbf{y})}{f(\mathbf{y}|\rho)} \mathrm{d}\mathbf{y}$ characterizes the mean information lost when $f(\mathbf{y}|\rho)$ is used to approximate $g(\mathbf{y})$. So, to fit the model $f(.|\rho)$ to g, one can look for the minimum of $\mathscr{K}(g;f(.|\rho))$, or of $\mathscr{B}(g;f(.|\rho)) \triangleq -\int g(\mathbf{y}) \ln f(\mathbf{y}|\rho) \mathrm{d}\mathbf{y}$, with respect to ρ.

Let $\mathbf{y}_1,\ldots,\mathbf{y}_J$ be J independent realizations of $\mathbf{y}$. The average negative log-likelihood of ρ, *i.e.* $\mathsf{NL}(\rho|\mathbf{y}_1,\ldots,\mathbf{y}_J) = -\frac{1}{J}\sum_{j=1}^{J} \ln f(\mathbf{y}_j|\rho)$, constitutes a finite-sample estimate of $\mathscr{B}(g;f(.|\rho))$. It can be computed even if g is unknown, and converges almost surely to $\mathscr{B}(g;f(.|\rho))$ as J tends to infinity. Its argmin, denoted by $\hat{\rho} = \arg\min_\rho \mathsf{NL}(\rho|\mathbf{y}_1,\ldots,\mathbf{y}_J)$, is merely the Maximum Likelihood Estimate (MLE) of ρ, *i.e.* $\hat{\rho} = \arg\max_\rho \prod_{j=1}^{J} f(\mathbf{y}_j|\rho)$. Under some regularity conditions, the MLE is known to be asymptotically efficient. So, $\mathsf{NL}(\rho|\mathbf{y}_1,\ldots,\mathbf{y}_J)$ when J tends to infinity, and thus $\mathscr{B}(g;f(.|\rho))$, are "good" criteria, in that they are most sensitive to a small deviation of $f(\mathbf{y}|\rho)$ from $g(\mathbf{y})$.

2) Information based Statistical Identification within Competing Models: Unfortunately, when handling several competing models, the MLE no longer provides a sound solution to statistical identification through the minimization of an estimate of $\mathscr{B}(g;f(.|\rho))$. This is the case when these models appear as different forms of $f(\mathbf{y}|\rho)$ or as a single $f(\mathbf{y}|\rho)$ but with different restrictions on the parameter vector ρ. Indeed, in this last case, the average negative log-likelihood $\mathsf{NL}(\rho|\mathbf{y}_1,\ldots,\mathbf{y}_J) = -\frac{1}{J}\sum_{j=1}^{J} \ln f(\mathbf{y}_j|\rho)$ systematically achieves a minimum when the optimization of ρ is performed over the highest-dimension admissible space.

In his seminal paper [10], Akaike derived the "A Information Criterion" $\mathsf{AIC}(\rho)$—now commonly referred to as the "Akaike Information Criterion"— as a fundamental basis to model selection. This criterion is defined so that $\frac{1}{J}\mathsf{AIC}(\rho)$ constitutes an estimate of $2\mathbb{E}[\mathscr{B}(g;f(.|\rho))]$, and writes as

$$\mathsf{AIC}(\rho) = -2\sum_{j=1}^{J} \ln f(\mathbf{y}_j|\rho) + 2k \tag{6}$$

where k is the number of free parameters in ρ. In the case of several competing models, the one leading to the minimum value of $\mathsf{AIC}(\rho)$ must be adopted, which will henceforth be named the MAICE, for "Minimum Akaike Information Criterion Estimate".

After Akaike's pioneering work, model selection was studied from quite different points of view. Schwarz first developed a Bayesian approach, which selects the most probable model *a posteriori* within a suitable family of candidates assigned with prior probabilities. Independently, Rissanen handled the models as a way to encode the observed data, and defined the optimal model as the "Minimum Description Length" (MDL), *i.e.* as the one which yields the minimum code length. In the large-sample limit, both approaches lead to select the model which minimizes the criterion

$$\mathsf{MDL}(\rho) = -\sum_{j=1}^{J} \ln f(\mathbf{y}_j|\rho) + \frac{1}{2}k\ln J. \tag{7}$$

For a fixed model, the minimum of (6) or (7) is the MLE $\hat{\rho} = \arg\min_\rho \mathsf{NL}(\rho|\mathbf{y}_1,\ldots,\mathbf{y}_J)$. Again, the additive corrective terms enable the comparison of competing models without erroneously selecting the one which entails the highest-dimension admissible parameter space, *i.e.* the maximum k.

B. Application to Source Detection: MAICE and MDL Source Number Estimates within a Farfield Narrowband MUSIC Scheme

1) Further assumptions and competing models definition: Recall that the number of active sources is the rank of the source covariance matrix $\underline{\mathbf{C}}_\mathbf{s} = \mathbb{E}[\mathbf{s}(t)\mathbf{s}^H(t)]$, which is entailed in the genuine array covariance matrix $\underline{\mathbf{C}}_\mathbf{x} = \mathbb{E}[\mathbf{x}(t)\mathbf{x}^H(t)]$ through (3). The aim is to detect $d = \mathrm{rank}(\underline{\mathbf{C}}_\mathbf{s})$ within a set of hypotheses from the sole knowledge of the finite-sample estimate $\underline{\tilde{\mathbf{C}}}_\mathbf{x}$ of $\underline{\mathbf{C}}_\mathbf{x}$. This can fit in the above generic framework by making some extra assumptions on the nature of the signals introduced in §II-B and by clearly defining the competing models. First, besides all the aforementioned assumptions, the sources complex envelopes and noise vectors are assumed to be Gaussian processes. Secondly, $\underline{\tilde{\mathbf{C}}}_\mathbf{x}$ is defined from J statistically independent samples $\mathbf{x}(t_1),\ldots,\mathbf{x}(t_J)$ of $\mathbf{x}$ as $\underline{\tilde{\mathbf{C}}}_\mathbf{x} = \frac{1}{J}\sum_{j=1}^{J} \mathbf{x}(t_j)\mathbf{x}^H(t_j)$. Last, to derive constructive decision algorithms from the data of $\underline{\tilde{\mathbf{C}}}_\mathbf{x}$, Wax and Kailath [11] suggest to consider $\underline{\tilde{\mathbf{C}}}_\mathbf{x}$ as a sample of the family

$$\underline{\mathbf{C}}_\mathbf{x}^{(d)} = \underline{\Psi}^{(d)} + \sigma^2 \underline{\mathbb{I}}_N, \tag{8}$$

with $\underline{\Psi}^{(d)}$ an unknown $N \times N$ Hermitian symmetric positive semidefinite matrix of rank $d \in \{1,\ldots,N-1\}$ and σ an unknown scalar. The parameter vector $\rho_\mathbf{x}^{(d)}$ then boils

down to the eigenvalues and signal space eigenvectors $\{\lambda_1,\ldots,\lambda_d,\sigma^2,\mathbf{U}_1,\ldots,\mathbf{U}_d\}$ of $\underline{\mathbf{C}}_{\mathbf{x}}^{(d)}$, because (8) also writes as $\underline{\mathbf{C}}_{\mathbf{x}}^{(d)}=\sum_{i=1}^{d}(\lambda_i-\sigma^2)\mathbf{U}_i\mathbf{U}_i^H+\sigma^2\mathbb{I}_N$.

2) Negative log-likelihoods and MLEs for fixed d: The negative log-likelihood $\mathsf{NL}(\rho_{\mathbf{x}}^{(d)}|\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}) \triangleq -\ln f(\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}|\rho_{\mathbf{x}}^{(d)}) = -\ln f(\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}|\underline{\mathbf{C}}_{\mathbf{x}}^{(d)})$, which is central to the definition of the AIC and MDL criteria, stems from the fact that $\mathbf{x}(t_1),\ldots,\mathbf{x}(t_J)$ are i.i.d. according to $\mathcal{CN}(\mathbf{x};\mathbf{0},\underline{\mathbf{C}}_{\mathbf{x}}^{(d)})$. Indeed,

$$\begin{aligned} f(\mathbf{x}(t_1),\ldots,\mathbf{x}(t_J)|\underline{\mathbf{C}}_{\mathbf{x}}^{(d)}) &= \\ &= \pi^{-NJ}(\det\underline{\mathbf{C}}_{\mathbf{x}}^{(d)})^{-J}\prod_{j=1}^{J}\exp-\left(\mathbf{x}(t_j)^H(\underline{\mathbf{C}}_{\mathbf{x}}^{(d)})^{-1}\mathbf{x}(t_j)\right) \\ &= \pi^{-NJ}(\det\underline{\mathbf{C}}_{\mathbf{x}}^{(d)})^{-J}\exp\left(-J\,\mathrm{trace}\left((\underline{\mathbf{C}}_{\mathbf{x}}^{(d)})^{-1}\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}\right)\right) \end{aligned} \quad (9)$$

can be viewed as $f(\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}|\underline{\mathbf{C}}_{\mathbf{x}}^{(d)})$, and the negative log-likelihood $\mathsf{NL}(\rho_{\mathbf{x}}^{(d)}|\tilde{\underline{\mathbf{C}}}_{\mathbf{x}})$ of $\rho_{\mathbf{x}}^{(d)}$ w.r.t. $\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}$ follows, with Z_1 a constant:

$$\mathsf{NL}(\rho_{\mathbf{x}}^{(d)}|\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}) = Z_1 + J\left(\ln\det\underline{\mathbf{C}}_{\mathbf{x}}^{(d)} + \mathrm{trace}\left((\underline{\mathbf{C}}_{\mathbf{x}}^{(d)})^{-1}\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}\right)\right). \quad (10)$$

After [11], the MLE $\hat{\rho}_{\mathbf{x}}^{(d)}=\arg\min_{\rho_{\mathbf{x}}^{(d)}}\mathsf{NL}(\rho_{\mathbf{x}}^{(d)}|\tilde{\underline{\mathbf{C}}}_{\mathbf{x}})$ for fixed d comes from the eigenvalues $l_1\geq l_2\geq\cdots\geq l_N$ and corresponding eigenvectors $\tilde{\mathbf{U}}_1,\ldots,\tilde{\mathbf{U}}_N$ of $\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}$ as

$$\hat{\lambda}_i=l_i,\ \hat{\mathbf{U}}_i=\tilde{\mathbf{U}}_i,\ i=1,\ldots,d;\quad \hat{\sigma}^2=\frac{1}{N-d}\sum_{i=d+1}^{N}l_i. \quad (11)$$

The minimum of $\mathsf{NL}(\rho_{\mathbf{x}}^{(d)}|\tilde{\underline{\mathbf{C}}}_{\mathbf{x}})$ follows, with Z_0 a constant:

$$\mathsf{NL}(\hat{\rho}_{\mathbf{x}}^{(d)}|\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}) = Z_0 - \ln\left(\frac{\prod_{i=d+1}^{N}l_i^{\frac{1}{N-d}}}{\frac{1}{N-d}\sum_{i=d+1}^{N}l_i}\right)^{J(N-d)}. \quad (12)$$

3) AIC and MDL criteria: The number $k^{(d)}$ of free entries in $\rho_{\mathbf{x}}^{(d)}$ amounts to

$$\begin{aligned} k^{(d)} &= \underbrace{d+1}_{\text{(I)}}\ \underbrace{+2dN}_{\text{(II)}}\ \underbrace{-2d}_{\text{(III)}}\ \underbrace{-2(d(d-1)/2)}_{\text{(IV)}} \\ &= d(2N-d)+1, \end{aligned} \quad (13)$$

where (I) is the maximum number of—real—distinct eigenvalues in the considered family of $\underline{\mathbf{C}}_{\mathbf{x}}^{(d)}$; (II) is the total number of coefficients of the—complex—entries of the signal space eigenvectors $\mathbf{U}_1,\ldots,\mathbf{U}_d$; (III) (resp. (IV)) account for the reduction of the degrees of freedom in $\rho_{\mathbf{x}}^{(d)}$ due to the normalization (resp. mutual orthogonality) of $\mathbf{U}_1,\ldots,\mathbf{U}_d$. Two strategies to the detection of narrowband sources can be deduced from (12)–(13). Given the eigenvalues $l_1\geq l_2\geq\cdots\geq l_N$ of $\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}$, they consist in minimizing either the AIC or MDL criteria below:

$$\begin{aligned} \mathsf{AIC}(d) &\triangleq \mathsf{AIC}(\hat{\rho}_{\mathbf{x}}^{(d)}) \\ &= -2J\ln\left(\frac{\prod_{i=d+1}^{N}l_i}{\left(\frac{1}{N-d}\sum_{i=d+1}^{N}l_i\right)^{(N-d)}}\right)+2d(2N-d) \end{aligned} \quad (14)$$

$$\begin{aligned} \mathsf{MDL}(d) &\triangleq \mathsf{MDL}(\hat{\rho}_{\mathbf{x}}^{(d)}) \\ &= -J\ln\left(\frac{\prod_{i=d+1}^{N}l_i}{\left(\frac{1}{N-d}\sum_{i=d+1}^{N}l_i\right)^{(N-d)}}\right)+\frac{1}{2}d(2N-d)\ln J. \end{aligned} \quad (15)$$

IV. EXTENSION TO BROADBAND FARFIELD SOURCES

Though the above developments can straightly cope with colored noise such that $\mathbb{E}[\mathbf{n}(t)\mathbf{n}^H(t)]=\sigma_{\mathbf{n}}^2\underline{\mathbf{C}}_{\mathbf{n}}$—by just replacing the eigendecomposition of $\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}$ by the generalized eigendecomposition of the matrix pencil $(\tilde{\underline{\mathbf{C}}}_{\mathbf{x}},\underline{\mathbf{C}}_{\mathbf{n}})$—an extension is needed to detect broadband sources.

A. Broadband Extensions of MUSIC

1) Basics: As thoroughly discussed in [12], two options can be taken when extending the MUSIC method to broadband signals. Both consist in applying a dedicated processing to their partitions onto B frequency "bins" k_b, $b=1,\ldots,B$, prior to turning the obtained information into a "composite pseudo-spectrum". In the first application of MUSIC to robot audition, [13] defined a broadband pseudo-spectrum as the average of separate pseudo-spectra independently computed on the B bins. Besides being computationally expensive—for it requires B eigendecompositions of $N\times N$ complex matrix pencils—this approach precludes any application of the above theory of information-theoretic detection. Contrarily, (14)–(15) extend to "coherent" schemes [14].

2) Coherent Broadband MUSIC: First, B full-rank "focalization matrices" $\underline{\mathbf{T}}(k_b)\in\mathbb{C}^{Q\times N}$, $b=1,\ldots,B$, are defined, with $D<Q\leq N$, so as to transform the array vector at any bin k_b into its value at a reference frequency k_0, *i.e.*, so that

$$\forall\theta,\ \underline{\mathbf{T}}(k_b)\mathbf{V}(\theta,k_b)=\underline{\mathbf{T}}(k_0)\mathbf{V}(\theta,k_0). \quad (16)$$

Then, summing the second order statistics of $\mathbf{Z}(k_b)\triangleq\underline{\mathbf{T}}(k_b)\mathbf{X}(k_b)\in\mathbb{C}^Q$ over all bins leads to the $Q\times Q$ "focalized array covariance matrix"

$$\underline{\Gamma}_{\mathbf{z}} \triangleq \sum_{b=1}^{B}\underline{\mathbf{T}}(k_b)\underline{\mathbf{C}}_{\mathbf{x}}(k_b)\underline{\mathbf{T}}^H(k_b) \quad (17)$$

$$= \underline{\mathbf{T}}(k_0)\mathbf{V}(\theta,k_0)\underline{\Gamma}_{\mathbf{s}}\mathbf{V}^H(\theta,k_0)\underline{\mathbf{T}}^H(k_0)+\sigma_{\mathbf{n}}^2\underline{\Gamma}_{\mathbf{n}}, \quad (18)$$

with $\underline{\Gamma}_{\mathbf{s}}\triangleq\sum_{b=1}^{B}\underline{\mathbf{C}}_S(k_b)$, and $\underline{\Gamma}_{\mathbf{n}}\triangleq\sum_{b=1}^{B}\underline{\mathbf{T}}(k_b)\underline{\mathbf{C}}_{\mathbf{n}}(k_b)\underline{\mathbf{T}}^H(k_b)$ the "focalized noise covariance matrix". Importantly, the generalized eigendecomposition $\{\lambda_i,\mathbf{U}_i\}_{i=1,\ldots,Q}$ of $(\underline{\Gamma}_{\mathbf{z}},\underline{\Gamma}_{\mathbf{n}})$ satisfies $\lambda_1\geq\ldots\geq\lambda_D>\lambda_{D+1}=\ldots=\lambda_Q=\sigma_{\mathbf{n}}^2$ and $\mathbf{V}^H(\theta,k_0)\underline{\mathbf{T}}^H(k_0)\mathbf{U}_{D+1}=\ldots=\mathbf{V}^H(\theta,k_0)\underline{\mathbf{T}}^H(k_0)\mathbf{U}_Q=\mathbf{0}^T$. From this generalized eigendecomposition, a single signal space can thus be defined—which is an approximately coherent combination of the signal spaces at all frequency bins—and the source azimuths can again be isolated as the maximum values of the pseudo-spectrum

$$h_{\text{broadband}}(\theta,k_0)\triangleq\frac{1}{\mathbf{V}^H(\theta,k_0)\underline{\mathbf{T}}^H(k_0)\underline{\Pi}_{\mathscr{N}}\underline{\mathbf{T}}(k_0)\mathbf{V}(\theta,k_0)}, \quad (19)$$

quite similar to (5) but with $\underline{\Pi}_{\mathscr{N}}\triangleq\sum_{i=D+1}^{Q}\mathbf{U}_i\mathbf{U}_i^H$.

Several important properties are in effect. First, the theoretical focalized covariance matrices are not available in practice, but are instead computed from estimates $\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}(k_b)$ and $\tilde{\underline{\mathbf{C}}}_{\mathbf{n}}(k_b)$. Secondly, the computational complexity of coherent broadband MUSIC is reduced, as the computation of the broadband pseudo-spectrum entails a single generalized decomposition of a $Q\times Q$ complex matrix pencil. Then,

reverberant environments entailing multipath propagation of several fully correlated—mirrored—sources can be dealt with, as soon as $\Gamma_{\mathbf{s}}$ has full-rank Q. Last, AIC and MDL criteria to source detection as introduced in §III-B can straightly fit into coherent broadband MUSIC.

B. MAICE *and* MDL *Source Number Estimates within a Broadband Beamspace MUSIC Scheme*

The above ideas are now instantiated into a strategy well-suited to robotics.

1) Selection of the Focalization Matrices: Each of the B focalization matrices $\underline{\mathbf{T}}(k_b) \in \mathbb{C}^{Q\times N}$, $b = 1,\dots,B$, can be built by stacking the row weights vectors of Q narrowband beamformers synthesized offline and matched to k_b. In other words, one defines $\underline{\mathbf{T}}(k_b) \triangleq \underline{\mathbf{W}}(k_b)^H$ where $\underline{\mathbf{W}}^H(k_b) = \left(\mathbf{w}_0(k_b) \mid \dots \mid \mathbf{w}_{Q-1}(k_b)\right)^H$ can be viewed as an operator from the N-dimensional microphones elementspace to a Q-dimensional output beamspace. The alignment property (16) is then equivalent to the invariance of the beampatterns $D_q(\boldsymbol{\theta},k_b) = \mathbf{W}_q^H(k_b)\mathbf{V}(\boldsymbol{\theta},k_b)$, $q = 0,\dots,Q-1$, across frequency bins $k_1,\dots,k_B$.

Following [15], an orthogonal beamspace processing structure can be obtained by setting the beampatterns to the spherical harmonics of increasing order $\mathscr{Y}_0(.),\dots,\mathscr{Y}_{Q-1}(.)$, *i.e.* by ensuring that $\forall q = 0,\dots,Q-1$, $\forall b = 1,\dots,B$, $\forall \boldsymbol{\theta}$, $D_q(\boldsymbol{\theta},k_b) = \mathscr{Y}_q(\boldsymbol{\theta})$. The corresponding QB row coefficients $\mathbf{W}_q^H(k_b) \in \mathbb{C}^{1\times N}$, $q = 0,\dots,Q-1$, $b = 1,\dots,B$, were synthesized for all bins thanks to the constructive method of [12].

2) Algorithm: The Algorithm 1 summarizes the prominent steps of the proposed detection strategy. It has been implemented on the EAR ("Embedded Audition for Robotics") integrated auditory sensor [16][17], made up with a uniform linear array of 8 microphones with even interspace $\frac{\lambda_{3\,\mathrm{kHz}}}{2} = 5.66\,\mathrm{cm}$, a fully programmable acquisition board, a FPGA processing unit, and USB communication. Prior to its hardcoding on the FPGA, an extended version of Algorithm 1 for $Q = 4$ has been implemented into a C/C++ library, so as to simultaneously detect and localize up to 3 broadband *nearfield* sources. To this aim, a loop on hypothesized ranges has been inserted between TIME_LOOP and BINS_LOOP, as was done in [12] for localization.

The suitability of the proposed strategy for robotics applications can be argued on several aspects. Indeed, many cumbersome computations are performed offline. The most involved online processing is undoubtedly the generalized eigendecomposition of the 4×4 matrix pencil $(\tilde{\underline{\Gamma}}_{\mathbf{z}}, \tilde{\underline{\Gamma}}_{\mathbf{n}})$—to be multiplied by the number of hypothesized range if nearfield sources detection is targeted. Nevertheless, from raw data sampled at 15 kHz, only 35% (resp. 57%) of a single Core of a DELL D630 laptop is required to compute and plot in real time @15 Hz pseudo-spectra for the hypothesized azimuths $0^\circ, 1^\circ, 2^\circ, \dots, 180^\circ$ and for 5 (resp. 50) hypothesized ranges.

V. EXPERIMENTS

The EAR sensor presented in §IV-B.2 is used. A first scenario concerns the detection and localization of a single broadband nearfield source in a silent but reverberating environment. The microphone array is fixed on a mast and oriented downwards, so as to privilege sounds emanating from a mobile phone on a table. The source true azimuth (73° w.r.t. endfire) and range (0.7 m) are determined thanks to a calibration chart laid on the table. Figure 1 reports coherent beamspace MUSIC pseudo-spectra. The number of sources is either assumed to be 1 (top) or detected using AIC (bottom). Though there is no ambient noise, pseudo-spectra may become inconsistent during soundtrack pauses (snapshots #12 and #36). Noticeably, the MAICE detects that no source is active. Furthermore, online source detection can lead to an improvement in subsequent MUSIC localization.

A second scenario takes place within a $5\mathrm{m}\times13\mathrm{m}$ room, with $\approx 0.45\,\mathrm{ms}$ reverberation time. A powerful air-conditioning system keeps humming and no sound absorbing material is used. The EAR sensor is mounted on a 1.5 m-high tripod, and placed 1.5 m parallel to a wall. Two loudspeakers $\mathsf{S}_1, \mathsf{S}_2$ are positioned on similar tripods at azimuths $62^\circ, 118^\circ$ by hand calibration—which is error prone up to some degrees. When a single loudspeaker utters a pure tone in $[600\,\mathrm{Hz}; 2.5\,\mathrm{kHz}]$ or a human voice, the detected number of sources is consistent with the soundtrack pauses and a—possibly asymmetric—bell-shaped histogram of the estimated azimuths along time is obtained. The mean estimate shows a bias less than $\pm10^\circ$ w.r.t. ground truth, and 95% of the estimates gather into an interval of less than 10°-width, see Figure 2. A case of two active sources is also shown.

Algorithm 1: Source Detection within a Broadband Beamspace MUSIC

OFFLINE, **do**
begin
- determine the B complex $Q\times N$ matrices $\underline{\mathbf{W}}(k_b)$, $b = 1,\dots,B$, by offline convex optimization as per [12]
- if the noise statistics $\underline{\mathbf{C}}_{\mathbf{n}}(k_b)$, $b = 1,\dots,B$, are not known, then "learn" an approximation $\tilde{\underline{\mathbf{C}}}_{\mathbf{n}}(k_b)$ from experimental data
- deduce the matrix $\tilde{\underline{\Gamma}}_{\mathbf{n}} = \sum_{b=1}^{B} \underline{\mathbf{W}}^H(k_b)\tilde{\underline{\mathbf{C}}}_{\mathbf{n}}(k_b)\underline{\mathbf{W}}(k_b)$

end

ONLINE, **do**
begin
for *each detection+localization time t;* /* TIME_LOOP */
do
for *each frequency bin k_b;* /* BINS_LOOP */
do
- compute the FFTs $\tilde{\mathbf{X}}_\tau(k_b)$ of $\mathbf{x}(.)$ over J non-overlapping groups of time snapshots indexed by $\tau_1,\dots,\tau_J \in]t-1,t]$
- deduce the estimates $\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}(k_b) \triangleq \frac{1}{J}\sum_{j=1}^{J} \mathbf{X}_{\tau_j}(k_b)\mathbf{X}_{\tau_j}^H(k_b)$ and $\tilde{\underline{\Gamma}}_{\mathbf{z}}(k_b) = \underline{\mathbf{W}}^H(k_b)\tilde{\underline{\mathbf{C}}}_{\mathbf{x}}(k_b)\underline{\mathbf{W}}(k_b)$ at time t

end
- deduce the focalized array covariance matrix $\tilde{\underline{\Gamma}}_{\mathbf{z}} = \sum_{b=1}^{B} \tilde{\underline{\Gamma}}_{\mathbf{z}}(k_b)$
- from the generalized eigenvalues $l_1 \geq \dots \geq l_Q$ and corresponding eigenvectors $\tilde{\mathbf{U}}_1,\dots,\tilde{\mathbf{U}}_Q$ of $(\tilde{\underline{\Gamma}}_{\mathbf{z}}, \tilde{\underline{\Gamma}}_{\mathbf{n}})$, **do**

begin
- compute the AIC or MDL criteria for each number d of hypothesized sources within $\{0,\dots,Q-1\}$ along formulae (14)–(15) with N replaced by Q therein
- detect the number of sources as $\hat{d} = \arg\min_{d\in\{0,\dots,Q-1\}}(\mathsf{AIC}(d) \text{ or } \mathsf{MDL}(d))$
- deduce the projector $\tilde{\underline{\Pi}}_{\mathscr{N}} \triangleq \sum_{i=\hat{d}+1}^{Q} \tilde{\mathbf{U}}_i\tilde{\mathbf{U}}_i^H$ and isolate the source bearings as the argmax of $\tilde{h}_{\text{broadband}}(\boldsymbol{\theta},k_0) \triangleq \frac{1}{\mathbf{V}^H(\boldsymbol{\theta},k_0)\underline{\mathbf{W}}(k_0)\tilde{\underline{\Pi}}_{\mathscr{N}}\underline{\mathbf{W}}^H(k_0)\mathbf{V}(\boldsymbol{\theta},k_0)}$

end

end

end

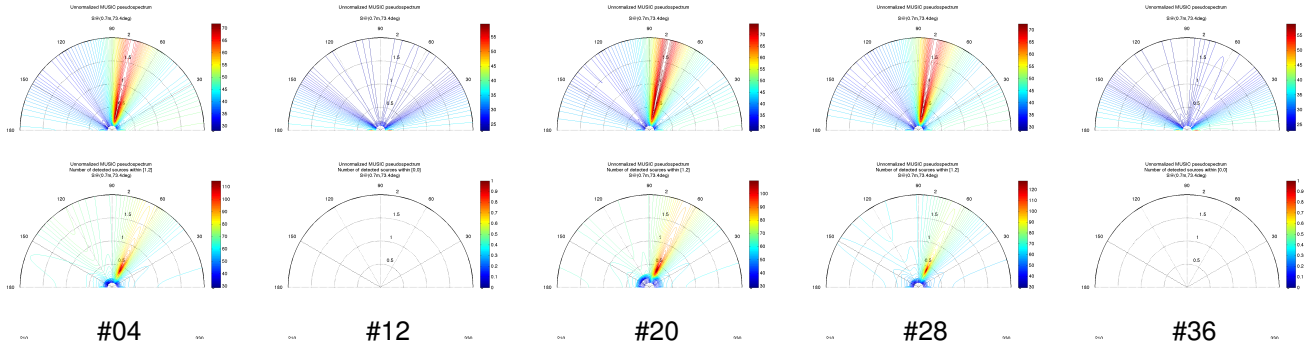

Fig. 1. Broadband beamspace MUSIC pseudo-spectra—in dB— *vs* (azimuth, range) under the assumption of a single source (top) or after AIC-based source detection (bottom). Pseudo-spectra iso-levels are drawn, the "hot" values tending to the peaks. #XX index the time snapshots.

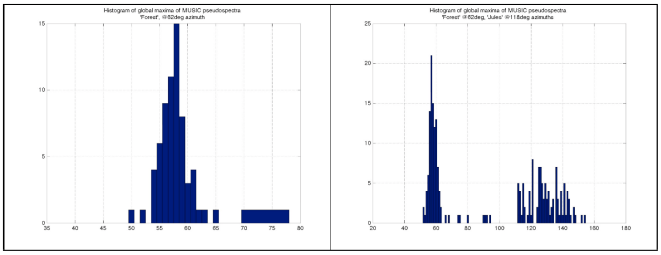

Fig. 2. Estimate histograms with AIC-based detection. S1@62° (left) and {S1@62°, S2@118°} (right) utter spoken messages.

VI. CONCLUSION AND OPEN PROBLEMS

A source detection method has been proposed, based on the introduction of AIC or MDL criteria within a coherent beamspace broadband MUSIC scheme. This method consists in minimizing over the hypothesized source numbers some mathematical expressions which only depend on the generalized eigenvalues of the matrix pencil made up with the focalized beamspace covariance and focalized noise covariance matrices. It is well-suited to robotics, because of its low complexity, its ability to cope with fully correlated sources, and its need of no prior threshold setting. Its implementation on the EAR sensor [17] has proved to perform well. The integration of the whole detection-localization scheme in the HARK software [18] is planned, as well as its hardcoding on the EAR FPGA unit.

As has been explained above, selecting either the AIC or MDL criterion depends on the preferred statement of the detection problem. Theoretically, MDL-based schemes almost surely detect the right number of sources in the large-sample limit ($J \longrightarrow +\infty$), while the MAICE tends to overestimate it, see [19] for more details.

To conclude, a trustworthy source detection method can enable a better exploitation of the outputs from localization algorithms. For instance, if localized sources are fewer in number than detected ones, then at least two sources emit from the same azimuth. Contrarily, if more sources are localized than their true number, then some estimated bearings come from false alarms and can be safely ignored.

REFERENCES

[1] R. Brückmann, A. Scheidig, C. Martin, and H.-M. Gross, "Integration of a sound source detection into a probabilistic-based multimodal approach for person detection and tracking," in *Autonome Mobile Systeme*. Springer Berlin Heidelberg, 2005, pp. 131–137.

[2] J. Valin, F. Michaud, B. Hadjou, and J. Rouat, "Localization of simultaneous moving sound sources for mobile robot using a frequency-domain steered beamformer approach," in *IROS'2004*, pp. 1033–1038.

[3] J.-S. Hu, C.-H. Yang, and C.-K. Wang, "Estimation of sound source number and directions under a multi-source environment," in *IEEE/RSJ IROS'2009*, Saint-Louis, MO, pp. 181–186.

[4] K. Nakamura, K. Nakadai, F. Asano, Y. Hasegawa, and H. Tsuhino, "Intelligent sound source localization for dynamic environments," in *IEEE/RSJ IROS'2009*, Saint-Louis, MO, pp. 664–669.

[5] C. Ishi, H. Ishiguro, and N. Hagita, "Evaluation of a MUSIC-based real-time sound localization of multiple sound sources in real noisy environments," in *IEEE/RSJ IROS'2009*, Saint-Louis, pp. 2027–2032.

[6] H. L. Van Trees, *Optimum Array Processing*, ser. Detection, Estimation, and Modulation Theory. John Wiley & Sons, Inc., 2002, vol. IV.

[7] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. on Antennas Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[8] B. Radich and K. Buckley, "The effect of source number underestimation on MUSIC location estimates," *IEEE Trans. on Signal Processing*, vol. 42, no. 1, pp. 233–235, Jan. 1994.

[9] D. Williams, "Detection: Determining the number of sources," in *The Digital Signal Processing Handbook*, Madisetti, V.K. and Williams D.B., Ed. Chapman & Hall, CRCnetBASE, 1999, ch. 67.

[10] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. on Automatic Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[11] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, Apr. 1985.

[12] S. Argentieri and P. Danès, "Broadband variations of the MUSIC high-resolution method for sound source localization in robotics," in *IEEE/RSJ IROS'2007*, San Diego, CA, pp. 2009–2014.

[13] F. Asano, H. Asoh, and T. Matsui, "Sound source localization and signal separation for office robot Jijo-2," in *IEEE/SICE/RSJ MFI'1999*, Taipei, Taiwan, pp. 243–248.

[14] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 33, no. 4, pp. 823–831, May 1985.

[15] D. Ward and T. Abhayapala, "Range and bearing estimation of wide-band sources using an orthogonal beamspace processing structure," in *IEEE ICASSP'2004*, Montreal, Canada, pp. 109–112.

[16] J. Bonnal, S. Argentieri, P. Danès, and J. Manhès, "Speaker localization and speech extraction with the EAR sensor," in *IEEE/RSJ IROS'2009*, Saint-Louis, MO, pp. 670–675.

[17] J. Bonnal, S. Argentieri, P. Danès, J. Manhès, P. Souères, M. Renaud, "The EAR project," *Jour. of the Robotics Society of Japan*, Jan. 2010.

[18] K. Nakadai, H. Okuno, H. Nakajima, Y. Hasegawa, and H. Tsuhino, "An open source software system for robot audition HARK and its evaluation," in *IEEE-RAS Humanoids'2008*, Daejeon, pp. 561–566.

[19] W. Xu and M. Kaveh, "Analysis of the performance and sensitivity of eigendecomposition-based detectors," *IEEE Trans. on Signal Processing*, vol. 38, pp. 1959–1971, Nov. 1990.