EB-ESPRIT: 2D LOCALIZATION OF MULTIPLE WIDEBAND ACOUSTIC SOURCES USING EIGEN-BEAMS

H. Teutsch, W. Kellermann

Multimedia Communications and Signal Processing
University of Erlangen-Nuremberg
Cauerstr. 7, 91058 Erlangen, Germany
{teutsch,wk}@LNT.de

ABSTRACT

This paper is concerned with the problem of localizing multiple wideband acoustic sources. In contrast to existing techniques, this method takes the physics of wave propagation into account. 2D wave fields are decomposed using cylindrical harmonics as basis functions by a circular array mounted into a rigid cylindrical baffle. The obtained wave field representation is then used to serve as a basis for high-resolution subspace beamforming methods, most notably ESPRIT. It is shown that acoustic source localization based on wave field decomposition has the potential to unambiguously localize multiple simultaneously active wideband sources in the array's full 360 degrees field-of-view.

1. INTRODUCTION

The problem of acoustic source localization has been an active area of research for more than a decade (see [1], [2], and references therein). Several applications depend on accurate position estimates of one or several simultaneously active sound sources, for example, tele-conferencing and surveillance systems.

Traditional sound source localization algorithms can be loosely divided into two major categories. The first group of algorithms is based on maximizing the output power of a steered beamformer. The second set of algorithms comprises two steps: First, the time delays of arrival (TDOA) between several microphone pairs are estimated. Second, these TDOAs are then used to estimate the source position using the information on the geometry of the setup. Many real-time systems based on these algorithms exist that prove satisfactory estimation performance.

One of the major shortcomings of these algorithms is the fact that the underlying signal model only allows a single sound source to be active at any given time in space.

A group of algorithms that resolves the simultaneous multiple source localization problem is based on high-resolution subspace techniques, such as MUSIC [3] and ESPRIT [4]. These techniques were originally developed for narrowband sources and were therefore not applicable to, e.g., speech signals. Although subspace techniques that are applicable to wideband signals exist (e.g. [5]), they do not seem to be considered as a viable alternative to the first two groups of algorithms in applications dealing with speech signals. This is partly due to their computational complexity which increases disproportionally with the number of microphones and due to the prerequisite of initial source location estimates.

All of the above mentioned algorithms have in common that they do not take the underlying physics of wave propagation in 2D or 3D space into account. In this paper, the physics of wave propagation in 2D space is considered for application to localization of a single and multiple wideband acoustic sources. 2D wave fields can be used as reasonable models for propagating acoustic sound fields in closed rooms where ceiling and floor reflections are sufficiently attenuated. A natural way of analyzing a 2D wave field is to decompose it into an orthogonal set of eigenfunctions of the acoustic wave equation in cylindrical coordinates, i.e. the cylindrical harmonics. The decomposed wavefield, in this paper denoted as 'eigen-space' [6], can be used to serve as a basis for many common subspace localization techniques, in particular ESPRIT. This idea is in principle similar to the familiar beam-space techniques which, before applying subspace localization algorithms, form several beams pointing in different directions in space and thus perform a reduction in computational complexity (see [7]). The decomposition into cylindrical harmonics can be achieved by utilizing circular microphone arrays that have the additional benefit of a full 360 degree field-of-view.

In this paper, a circular microphone array mounted into a rigid cylindrical baffle is examined. The decomposition is described in Section 2, the eigen-beam ESPRIT (EB-ESPRIT) algorithm is presented in Section 3, and its performance is evaluated in Section 4.

2. CIRCULAR ARRAYS IN EIGEN-SPACE

Figure 1 depicts the geometric model under consideration where a planar wave front impinges on a circular aperture of radius R that is mounted into a rigid cylindrical baffle. Due to the fact that any circular aperture has control only over the horizontal component of a wave field, $\theta_i = \pi/2$ is assumed throughout this paper.

The Fourier series expansion of a planar wave field due to a single far-field source, expressed in polar coordinates, is

$$B(kr, \phi) = \sum_{n=-\infty}^{\infty} j^n A_n(kr) e^{jn\phi}, \tag{1}$$

where $k = |\boldsymbol{k}|$ is the wavenumber and $j^2 = -1$. Depending on the applicable boundary conditions [8], it follows that

$$A_n(kr) = \begin{cases} J_n(kr), & \text{w/o baffle} \\ J_n(kr) - \dfrac{J_n'(kR)}{H_n^{(1)\prime}(kR)} H_n^{(1)}(kr), & \text{rigid cylinder} \end{cases} \tag{2}$$

where $J_n(\cdot)$ is the Bessel function of the first kind of order n and $H_n^{(1)}(\cdot)$ is the n-th order Hankel function of the first kind. The

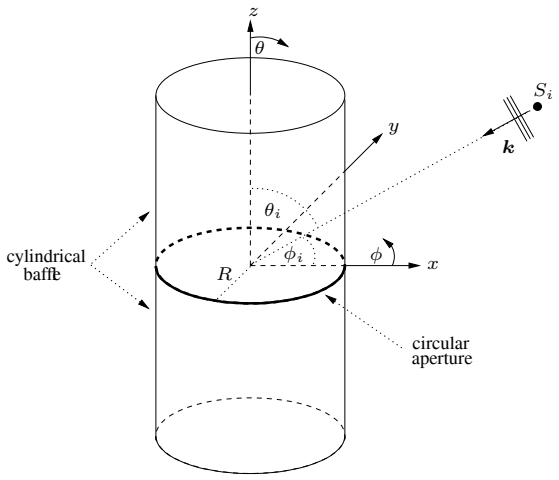

Fig. 1. *Geometric model*

prime denotes the derivative with respect to the argument. By considering $A_n(kr) = J_n(kr) - [J_n(kR) \cdot H_n^{(1)}(kr)]/H_n^{(1)}(kR)$, cylinders representing 'pressure-release' boundaries [9] can also be modeled.

If the circular aperture is not mounted into a cylindrical baffle (1) evaluates to $e^{jkr\cos\phi} = \sum_{n=-\infty}^{\infty} j^n J_n(kr) e^{jn\phi}$, the standard free-field expansion for plane waves [9].

The response of a circular aperture of $r = R$ due to a single plane wave impinging from angle ϕ_i can be written as,

$$F(k,\phi_i) = \frac{1}{2\pi}\int_0^{2\pi} \underbrace{\sum_{n=-\infty}^{\infty} j^n A_n(kR) e^{jn(\phi-\phi_i)}}_{\text{total wave field on aperture}} \quad \underbrace{\varrho(k,R,\phi)}_{\text{aperture illum.}} \quad d\phi. \tag{3}$$

The aperture illumination, $\varrho(R,k,\phi)$, in general, can be regarded as a frequency-dependent weighting function for an infinitesimal segment, $d\phi$, of the aperture. By choosing $\varrho(R,k,\phi) = e^{-jm\phi}$, $m \in \mathbb{N}$, it can be shown that, by orthogonality of the exponential functions, Eq. (3) reduces to

$$F_n(k,\phi_i) = j^n \, A_n(kR) \, e^{-jn\phi_i}. \tag{4}$$

In other words, by simply applying a Fourier transform with respect to ϕ to the response of a circular aperture one obtains the so-called cylindrical harmonics as defined in (4). By utilizing the fast Fourier transform (FFT), an efficient implementation of the decomposition into cylindrical harmonics can be obtained. Multiple incoming plane waves can be taken into account by superposition of the respective individual harmonics.

Figure 2 shows the frequency-dependent component of the cylindrical harmonics, i.e. $A_n(kR)$. Both apertures exhibit a high-pass character up to $kR \sim 1$ with a slope of $6n$ dB/octave. The advantage of mounting a circular aperture into a rigid baffle now becomes clear. The zeros in the amplitude response of the individual harmonics caused by the Bessel functions (left figure) are compensated by the additional components present in the setup employing the rigid cylindrical baffle (right figure). This fact makes this arrangement useful for a wider frequency range.

Figure 3 shows the angular-dependent component of the cylindrical harmonics, i.e. $e^{-jn\phi_i}$. As can be seen, these harmonics

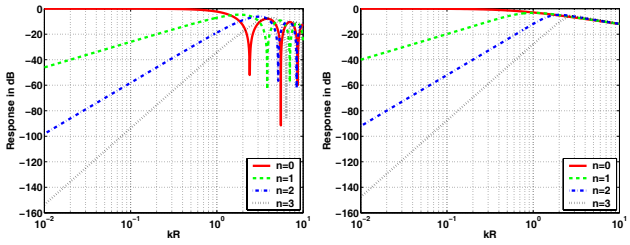

Fig. 2. *Frequency response (left: w/o baffle, right: rigid baffle)*

correspond to multipoles, i.e. monopole, dipole, quadrupole, etc, which are mutually *orthogonal*. As will be shown in Section 3, these multipoles can be used instead of the individual microphone responses for subspace-based wideband source localization.

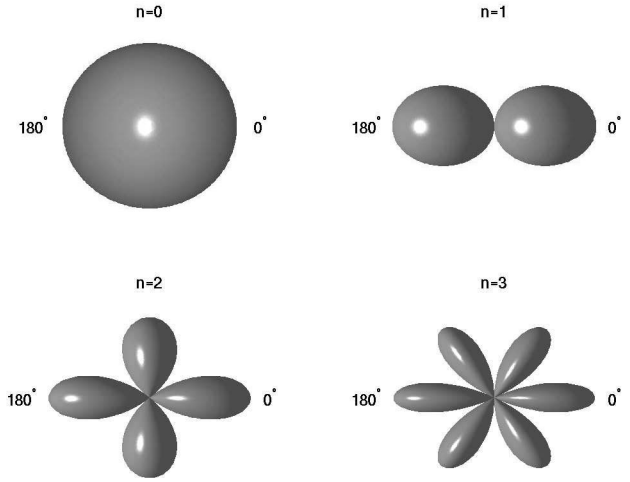

Fig. 3. *Angular response with respect to ϕ*

So far, only circular apertures have been considered. For actual implementations, the aperture has to be sampled at discrete points in space, i.e. microphones. Therefore (3) needs to be discretized to yield $\hat{F}_n(k,\phi_i)$. It can be shown that $\hat{F}_n(k,\phi_i) = F_n(k,\phi_i) + \mathcal{E}(n_a,k,\phi_i)$, where $\mathcal{E}(n_a,k,\phi_i)$ is an additional term due to modal aliasing which basically results in several modes of order $n_a > n$ leaking into mode n. This error can be controlled, although not eliminated, by appropriately choosing the number of microphones M, the radius R, and the frequency range of operation. Further details can be found in [10].

3. ESPRIT ALGORITHM FOR SOURCE LOCALIZATION

In order to be able to apply the ESPRIT algorithm, in general, the frequency-dependence of the individual harmonics must be compensated, so that

$$G_n(k,\phi_i) = \frac{\hat{F}_n(k,\phi_i)}{j^n A_n(kR)} \approx e^{-jn\phi_i}, \tag{5}$$

where it is assumed that $\mathcal{E}(n_a,k,\phi_i)$ is sufficiently small. Hence, (5) can be represented by a frequency-independent quantity $\tilde{G}_n(\phi_i)$. It is further assumed that K sources impinge on the circular array.

Therefore, the output at time t can be written as

$$\boldsymbol{y}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}(t), \tag{6}$$

where $\boldsymbol{A} = [\boldsymbol{G}(\phi_1) \mid \ldots \mid \boldsymbol{G}(\phi_K)]$ is a modal array matrix with $\boldsymbol{G}(\phi_i) = [\tilde{G}_{-N}(\phi_i) \ldots \tilde{G}_N(\phi_i)]^T$ and where the source vector is defined as $\boldsymbol{s}(t) = [s_1(t) \ldots s_K(t)]^T$. $\boldsymbol{n}(t)$ denotes the noise vector at time t and $(\cdot)^T$ denotes transposition. N is the highest mode of the decomposition to be taken into consideration, e.g. in Figs. 2 and 3 it follows that $N = 3$. Therefore, a total of $O = 2 \cdot N + 1$ independent harmonics can be used for further processing. It can be shown that, theoretically, it is then possible to localize $K = N$ sources.

The covariance matrix of the modal array response is

$$\boldsymbol{R}_{yy} = E\{\boldsymbol{y}(t)\boldsymbol{y}^H(t)\}, \tag{7}$$

where $E\{\cdot\}$ denotes the expectation operator and $(\cdot)^H$ the Hermitian operator. Assuming zero-mean spatially and temporally white Gaussian noise, (7) can be expressed as

$$\boldsymbol{R}_{yy} = \boldsymbol{A}\boldsymbol{R}_{ss}\boldsymbol{A}^H + \sigma^2\boldsymbol{I}, \tag{8}$$

where $\boldsymbol{R}_{ss}$ is the signal covariance matrix, $\boldsymbol{I}$ denotes the identity matrix, and σ^2 is the noise variance. It can be shown that the eigenvectors of $\boldsymbol{R}_{yy}$, corresponding to the K largest eigenvalues, form the signal subspace $\boldsymbol{E}_S$. These eigenvectors are linear combinations of the modal array vectors of $\boldsymbol{A}$ corresponding to the K signal sources.

Now, the standard sensor-space ESPRIT [4] can be directly applied to eigen-space after replacing the notion of an individual microphone of a standard linear microphone array with an individual harmonic of order n (see Fig. 4). It is assumed that the

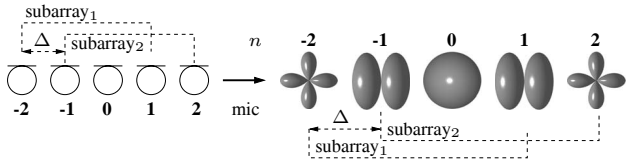

Fig. 4. *Transition from sensor-space to eigen-space for* $N = 2$ *and* $\Delta = 1$

first element in the original sensor-space (eigen-space) array is the first element in the first subarray (sub-modal array) and that the $(\Delta + 1)$th element in the original array is the first element in the second subarray. Note that Δ, therefore, does *not* correspond to a physical shift in either array configuration. The motivation for this transition can be justified by realizing that the modal array matrix $\boldsymbol{A}$ is Vandermonde, just like the array matrix of a linear array. A significant difference, however, is the fact that $\boldsymbol{A}$ is *frequency-independent*. This observation is the reason why no focusing matrices are needed for aligning the individual narrowband signal subspaces of wideband signals.

Following the sensor-space ESPRIT algorithm [4], two subarrays of length $\hat{M}$ and displacement $\Delta \in \mathbb{N}$, cf. Fig. 4, are chosen whose subarray (sub-modal array) matrices satisfy the invariance relation

$$\boldsymbol{A}_2 = \boldsymbol{A}_1\boldsymbol{\Phi}, \tag{9}$$

where $\boldsymbol{\Phi} = \mathrm{diag}\{e^{-j\Delta\phi_1}, \ldots, e^{-j\Delta\phi_K}\}$ and

$$\boldsymbol{A}_1 = [\tilde{\boldsymbol{K}}_s \mid \boldsymbol{0}] \cdot \boldsymbol{A}, \quad \boldsymbol{A}_2 = [\boldsymbol{0} \mid \tilde{\boldsymbol{K}}_s] \cdot \boldsymbol{A}. \tag{10}$$

$\tilde{\boldsymbol{K}}_s$ is an $\hat{M} \times \hat{M}$ identity matrix and $\boldsymbol{0}$ is a $\hat{M} \times \Delta$ zero matrix (e.g. 4×1 in Fig. 4). Since the columns of $\boldsymbol{A}$ span the signal subspace, $\boldsymbol{E}_S$, it holds that

$$\boldsymbol{E}_S = \boldsymbol{A}\boldsymbol{T}, \tag{11}$$

where $\boldsymbol{T}$ is a non-singular matrix. Therefore, the subspaces of the two subarrays can be defined as

$$\boldsymbol{E}_{S1} = \boldsymbol{A}_1\boldsymbol{T}, \tag{12}$$

$$\boldsymbol{E}_{S2} = \boldsymbol{A}_2\boldsymbol{T} = \boldsymbol{A}_1\boldsymbol{\Phi}\boldsymbol{T}. \tag{13}$$

By combining (12) and (13) one obtains

$$\boldsymbol{E}_{S2} = \boldsymbol{E}_{S1}\boldsymbol{\Psi}, \tag{14}$$

where $\boldsymbol{\Psi} = \boldsymbol{T}^{-1}\boldsymbol{\Phi}\boldsymbol{T}$. $\boldsymbol{\Psi}$ can then be obtained from (14) by applying a standard least-squares or total least-squares solver [11]. By realizing that the eigenvalues of $\boldsymbol{\Psi}$ are the diagonal elements of $\boldsymbol{\Phi}$ the locations of the sources can be estimated. Note that in a real system all equalities in (11),(12), and (13) must be replaced by approximate equalities since they must be estimated from the observed covariance matrix, $\hat{\boldsymbol{R}}_{yy}$, which is subject to measurement errors.

Note also, although not shown here, that for a single source an algorithm can be formulated that does not require the frequency compensation of (5). This fact has the advantage that uncorrelated noise does not get amplified at low frequencies by lowpass-like equalization filters exhibiting a slope of $6n$ dB/octave, cf. Fig 2.

4. EVALUATION

In order to show the ability of EB-ESPRIT to localize one and two sources, two sets of performance evaluations were undertaken with $M = 10$, $N = 3$, and $K = 1, 2$. The first set of evaluations was based entirely on computer simulations, i.e. the microphone array mounted into a cylindrical baffle can therefore be considered as perfectly modeled. This set of evaluations is in the following denoted as 'simulated'.

Fig. 5. *Photograph of the ten-element microphone array system with* $R = 0.04\,m$

For the second set of evaluations, denoted as 'measured', an actual circular microphone array ($R = 0.04\,\mathrm{m}$) comprising ten

omnidirectional microphones mounted into a rigid cylindrical baffle has been realized (see Fig. 5). Impulse responses from 48 loudspeakers surrounding the microphone array at a distance of 1.5m to each microphone were measured. The measurements were performed in a room with $T_{60} \approx 250$ ms. All further processing was done offline for the sake of clarity of presentation.

Figure 6 shows the performance of the system due to a four-second signal segment that contains single male unreverberated speech for $kR \in [0.22 \ldots 2.2]$, which is equivalent to a frequency-range of $f \in [300 \ldots 3000]$ Hz. The evaluation involved the following steps. First the signal coming from $\phi_0 = 35°$ was divided into data blocks of 100 ms length. Then each block of data was either modified by additive noise, resulting in an SNR of 15 dB (simulated), or convolved with impulse responses from the respective loudspeaker position to the individual microphones (measured). This procedure was then repeated 11 times so that $\phi = \phi_0 + l \cdot \phi_a$ where $\phi_a = 30°$ and $l = [1, \ldots, 11]$. The resulting position estimates, $\hat{\phi}$, are shown in the respective superimposed plots of Fig. 6. Although not immediately obvious from the figures, the maximum variance was found to be less than $0.02°$ (simulated) and $0.9°$ (measured), respectively.

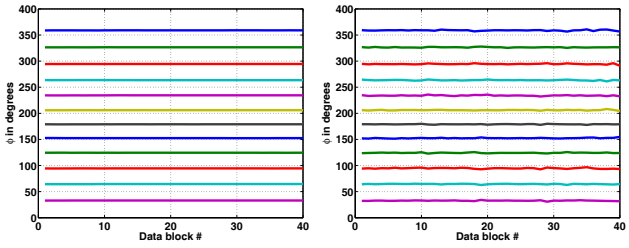

Fig. 6. *Single plane wave (overlaid illustration, left: simulated, right: measured)*

Figure 7 shows the position estimates, $\hat{\phi}$, for two simultaneously impinging wave fronts with $\phi = [15° \ 80°]$ and signal powers of -6 dB and 0 dB subject to an SNR of 15 dB, respectively. Here, the frequency range was reduced to $kR \in [0.7 \ldots 2.2]$ which is equivalent to $f \in [1000 \ldots 3000]$ Hz in order to to minimize the effects introduced by the noise amplification properties of (5). In

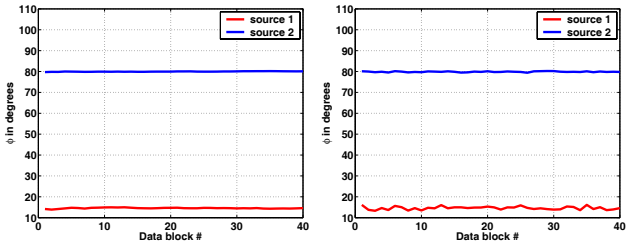

Fig. 7. *Two plane waves (left: simulated, right: measured)*

these experiments, uncorrelated white noise was chosen as the excitation signals since the problem of how to *reliably* estimate the number of active speech signals is yet unsolved. Note that in the derivation of the EB-ESPRIT algorithm it was assumed that the number of active sources, i.e. the number of principal eigenvectors of the covariance matrix, is known. The results of the localization algorithm in the right-hand side of Fig. 7 show an encouraging maximum variance for the weaker source of $0.6°$.

The above derived EB-ESPRIT has been implemented as a real-time demonstrator running at about 50% CPU usage of a dual-processor Pentium4 1 GHz processor under the Linux operating system. Thereby, employing the array shown in Fig. 5, the results obtained through offline calculations have been verified under real (non-stationary) conditions.

5. CONCLUSIONS

An eigen-space version of ESPRIT has been derived that is based on the decomposition of a 2D wave field into an orthogonal set of eigensolutions to the acoustic wave equation, i.e. the cylindrical harmonics. These harmonics can be obtained by utilizing circular microphone arrays with or without mounting them into a rigid cylindrical baffle. This formalism is applicable to 3D wave fields as well. In this case, spherical microphone arrays are needed to decompose the wave field into *spherical harmonics* [6]. Localization will then also yield an estimate of the source's elevation [12]. Future work will include a methodology for estimating the number of active speech signals in a wave field.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gary W. Elko for many fruitful discussions.

6. REFERENCES

[1] M. Brandstein and D. Ward, Eds., *Microphone Arrays – Signal Processing Techniques and Applications*, Springer, 2001.

[2] Y. Huang and J. Benesty, Eds., *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Kluwer, 2004.

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

[4] R. Roy and T. Kailath, "ESPRIT – estimation of signal parameters via rotational invariance techniques," *IEEE Trans. on Acoust., Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.

[5] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wideband sources," *IEEE Trans. on Acoust., Speech, and Signal Processing*, vol. ASSP-33, pp. 823–831, August 1985.

[6] J. Meyer and G.W. Elko, "A highly scalable spherical microphone array based on a orthonormal decomposition of the soundfi eld," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Orlando, Florida, USA, May 2002, IEEE, pp. II–1781–II–1784.

[7] G. Xu, S.D. Silverstein, R. Roy, and T. Kailath, "Beamspace ESPRIT," *IEEE Trans. on Signal Processing*, vol. 42, no. 2, pp. 349–356, February 1994.

[8] M.C. Junger and D. Feit, *Sound, Structures, and Their Interaction*, Acoustical Society of America, 1993.

[9] E.G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography*, Academic Press, London, 1999.

[10] C.P. Mathews and M.D. Zoltowski, "Eigenstructure techniques for 2-D angle estimation with uniform circular arrays," *IEEE Trans. on Signal Processing*, vol. 42, no. 9, pp. 2395–2407, September 1994.

[11] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 2 edition, 1989.

[12] H. Teutsch and W. Kellermann, "Eigen-beam processing for direction-of-arrival estimation using spherical apertures," in *Proc. Joint Workshop on Hands-Free Communication and Microphone Arrays*, Piscataway, NJ, March 2005.