Modal Analysis Based Beamforming for Nearfield or Farfield Speaker Localization in Robotics

Sylvain Argentieri†, Patrick Danès†‡ and Philippe Souères†
{sargenti,danes,soueres}@laas.fr
†LAAS - CNRS
7 avenue du Colonel Roche, 31077 Toulouse, France
‡Université de Toulouse (UPS)
118 route de Narbonne, 31062 Toulouse, France

Abstract— **This paper describes a broadband beampattern synthesis method for sound source localization in the nearfield or in the farfield of a mobile robot, with a small-size linear array. The method is based on the theory of Modal Analysis and involves an original convex optimization procedure which benefits from the Parseval relation. The optimized beampattern is obtained by numerically minimizing the worst-case error between the modal coefficients of the array response and those of the reference beampattern, up to a finite rank of the series expansion, over a frequency grid. Simulations illustrate the analytical development.**

I. Introduction

To endow a mobile robot with the capability to locate sound sources in real-time, in a variable and noisy environment, one needs to tackle difficult problems for which classical techniques of acoustics do not directly apply. In most part of applications, the signal of interest is broadband and the distance to the source may vary from a few tens of centimeters to several meters. Acousticians usually distinguish the nearfield neighborhood of the source, inside which the soundwave is clearly spherical, from the farfield region where it can be considered as planar. However, though the dependence on distance is fundamental, very few localization methods allow to consider it explicitly.

For a large part of methods that involve a pair of microphones, and which are based on the estimation of the Interaural Phase and Intensity Differences (IPD and IID), experiments report satisfactory results at a specific distance from the source, usually about one meter. In these works, the localization accuracy worsens rapidly with the distance. Considering the human auditory system, some authors proposed to determine the robot's Head Related Transfer Function (HRTF) to estimate the IPD and IID between two microphones [1] [2]. An azimuthal speaker localization method based on TDOA and cross-correlation techniques provided accurate results at short distance [3]. A 3D speaker tracking method using TDOA and a multiple model adaptive estimator based on Kalman filtering was described in [4]; here again the localization was satisfactory at 1 meter from the source but deteriorated rapidly with the distance. Besides, beamforming techniques have also been applied to robotics with different kinds of microphone arrays. A localization method using a frequency domain delay-sum beamformer with a cubic array was developed in [5]. Delay-sum beamforming and frequency band selection techniques were used for sound localization and separation in [6] with a three-ring microphone array. In order to maintain a constant localization accuracy over the whole frequency band, convex optimization techniques were used in [7] with a small-size linear array. However, these applications of beamforming techniques do not explicitly address the question of distance, and make the implicit hypothesis that the soundwave is planar. As a consequence, such methods turn out to be inappropriate to localize broadband signals in the nearfield.

In this paper, we describe a broadband beampattern synthesis method, based on the theory of Modal Analysis and involving an original convex optimization procedure. Modal Analysis allows to express any beampattern, which comes as a function of the distance, the direction and the frequency, as an infinite series of analytic functions depending on a set of Modal coefficients [8]. The optimized beampattern is obtained by numerically minimizing the worst case error between the Modal coefficients of the antenna response and those of the reference beampattern, up to a certain finite rank, and on a frequency grid. The paper is organized as follows. Having given a brief account on farfield broadband beamforming in section II, some elements of the theory of Modal Analysis are introduced in section III. Using this representation, we show that a beamformer designed under the farfield assumption is inappropriate in the nearfield, especially at low frequencies. The proposed Array Pattern Synthesis (APS) method is presented in section IV. Simulation results are given to illustrate the analytical reasoning. Finally, a discussion concerning this APS method and a comparison with previous results based on convex optimization [7] are given in section V.

II. Farfield and nearfield beamforming

This section recalls some generalities on sound source localization from an array of microphones through beamforming. Recent works have shown that this approach can deal with broadband signals, and allows to cope with specific robotics constraints (small-size embeddable arrays, real-time localization, …). However, most beamforming applications to robotics assume that the sound source emits in the farfield, i.e. the soundwave impinging on the sensor is planar. Considering a previous contribution [7], we show in the last part of this section that array patterns designed under the farfield hypothesis show a significant distortion in the nearfield.

A. Generalities on beamforming

In the following, f terms the temporal frequency of a pointwise sound source, $\lambda = \frac{c}{f}$ is its associated wavelength –with

$c = 340m.s^{-1}$ the speed of sound– and $k = \frac{2\pi f}{c}$ terms its wavenumber. Consider a rectilinear array of N microphones, positioned at abscissae $z_1, \ldots, z_n$ on the z-axis, and a sound source emitting at point M (Figure 1). The source is said to be in the *farfield* (resp. in the *nearfield*) if the spatial sampling of the wavefield performed by the sensor array is such that the wavefronts can be assumed locally *planar* (resp. must be modelled as *spherical*). As illustrated later, this farfield/nearfield assumption depends not only on the distance r from M to the origin 0, but also on the source frequency k.

Classically, the farfield assumption is often used to deal with sound source localization. In this context, the linear array farfield response $D^{\infty}(\theta, k)$, called *array pattern* or *beampattern*, is a function of the frequency k and direction θ of the incoming waves. It can be tuned up to a suitable design through *beamforming*, whose objective is to electronically polarize the array so as to make it sensitive to signals coming from a specific direction of interest θ_0 and covering a given range of frequencies. The same holds for nearfield beamforming, in which case the array beampattern $D(r, \theta, k)$ explicitly depends on the distance r. A direct

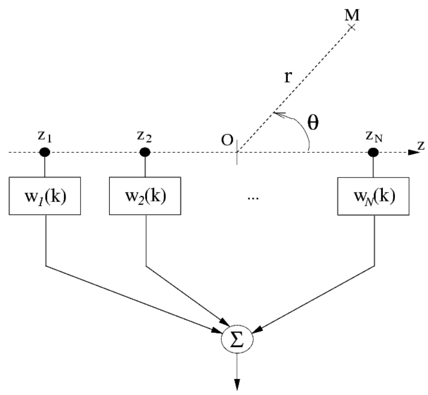
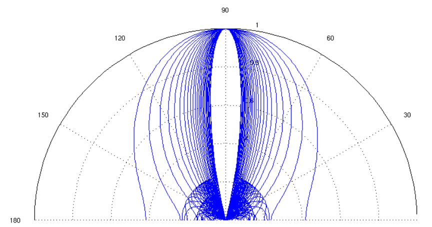


Fig. 1. Typical farfield ($r \to \infty$) or nearfield broadband beamformer

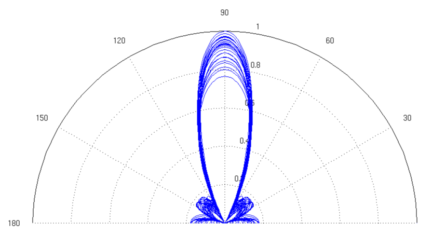
method for source angular localization immediately follows. It consists in making successive hypotheses on the *Direction Of Arrival* (DOA) by scanning every direction θ_0. For each assumption, the beamformer's output energy is computed over a temporal sliding window, leading to an *acoustic energy map* that is expected to be maximum at the actual source DOA. For general waves like voice, broadband beamformers must be used, which often associate a separate filter to each microphone (Figure 1). This type of beamforming is also called *filter-sum beamforming*.

B. Farfield beamforming

To make the array sensitive to one spatial DOA θ_0 along a whole range of frequencies, one straight strategy consists in compensating the time delays due to propagation between the microphones. It has been shown in [9] that the consequent beamformer –henceforth called "classical"– shows a main lobe centered on θ_0, whose width increases at low frequencies. The beampattern of an $N = 8$-microphone array with even interspace $\frac{1}{2}\lambda_{3kHz} = 5.66cm$ is shown in Figure 2(a) for $\theta_0 = 90°$ and frequencies ranging between $400Hz$ and $3kHz$. Clearly, the important width of the main lobe leads to a poor focalization in the listening direction. Nevertheless, this classical beamformer has been widely used in robotics applications [5] [6]. More generally, beamforming techniques



(a) Farfield Classical Filter-Sum Beamformer



(b) Farfield Optimized Filter-Sum Beamformer

Fig. 2. Antenna pattern for k $\in K$, with one curve per frequency. Large main lobes correspond to low frequencies.

for broadband signals often select for each $w_n(k)$ a separate Q^{th}-order Finite Impulse Response (FIR) filter. In other words, the frequency response of the filter associated to the n^{th} microphone, $n = 1, \ldots, N$, reads as

$$w_n(k) = \sum_{q=0}^{Q} w_{nq} e^{-ikcqT_e}, \tag{1}$$

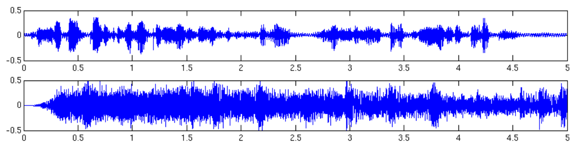
with w_{nq} its q^{th} coefficient and T_e the sampling period. Consequently, the array pattern $D^{\infty}(\theta, k)$ in the farfield can be written [10]

$$D^{\infty}(\theta, k) = \sum_{n=1}^{N} \sum_{q=0}^{Q} w_{nq}\ e^{-ikcqT_e}\ e^{ikz_n \cos\theta}. \tag{2}$$

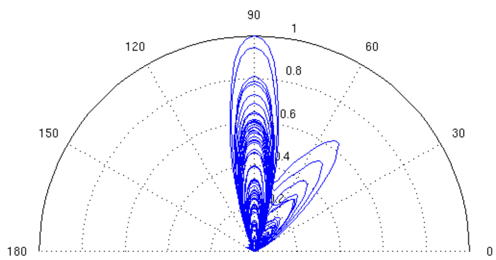
All the difficulty lies in the selection of an Array Pattern Synthesis (APS) method for computing the $N \times Q$ coefficients. Under the farfield assumption, we recently proposed an APS method based on convex optimization [7]. Therein, we claimed that a better array response can be got compared to classical farfield broadband beamforming techniques previously used in robotics. A design example, on the frequency band $[0.5 : 0.1 : 3]$kHz and $\Theta = [0 : 2 : 180]°$, is plotted[1] on Figure 2(b). The optimized result shows that the main lobe width is nearly constant over the whole range of frequencies: consequently, the speaker localization should enjoy a better resolution. This fact can be easily confirmed by simulating the localization process which computes an acoustic energy map of the environment. This localization map is obtained by

[1]MATLAB-like conventions are used to represent linearly spaced vectors, i.e. $V = $ [begin : step : end].

simulating the output energy of each optimized beamformer over a fixed-width sliding temporal window (see [7] for more details on the algorithm). In the sequel, we set $T_{obs} = 1000T_e$, with $f_e = \frac{1}{T_e} = 15kHz$. Let's first consider two sources, emitting from the direction $\theta_{voice} = 55°$ and $\theta_{noise} = 90°$. The first one is a male voice, the second one is a music extract. The two signals are plotted in Figure 3(a). Figure 3(b) shows the computed energy maps when the two sources are far enough for the planar wavefront assumption to hold. In this



(a) Voice and music signals



(b) Farfield energy map

Fig. 3. Normalized energy map computed by a farfield beamformer over successive temporal windows (one curve per window) for farfield sound sources.

case, the optimized strategy leads to an efficient localization, the energy map showing two prominent lobes along the DOA's of the two sources. However, this nice property is only guaranteed in the farfield, while in many cases the small distance between the speaker and the array, together with the low frequency content of his voice, require to explicitly handle the wavefronts sphericity. Consequently, the performance of such a beamformer must be analyzed when the sound source is in the nearfield.

C. Influence of nearfield sources on a farfield beamformer

When dealing with the nearfield case, the distance r from the origin 0 to the sound source must explicitly be taken into account. This leads to the general array response [10]

$$D(r,\theta,k) = \sum_{n=1}^{N}\sum_{q=0}^{Q} \frac{w_{nq}\ r}{d_n(r,\theta)}\ e^{-ikcqT_e}\ e^{ik(d_n(r,\theta)-r)}, \quad (3)$$

where $d_n(r,\theta)$ terms the distance from the source to the n^{th} sensor. Notice that the general beampattern expression (3) turns into Equation (2) when $r \to \infty$. This general expression can be used to evaluate the performance of a beampattern designed under the farfield assumption, but used in the nearfield. For instance, the farfield beampattern plotted on Figure 2(b) turns into the one plotted in Figure 4 when considering the distance $r = 0.8m$. The performances of the array at short r are clearly modified: the main lobe widens at low frequencies, while high frequencies seem unchanged. In fact,

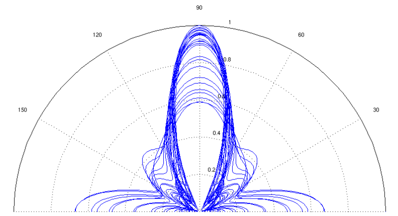


Fig. 4. Farfield Optimized Filter-Sum Beamformer in the nearfield ($r = 0.8m$). Large main lobes correspond to low frequencies.

low frequencies must be explicitly modeled in the nearfield of the array whereas high frequencies, *at the same distance*, can be considered in the farfield. As the voice power is mainly located at low frequencies, this pattern alteration can significantly worsen the resolution of the speaker localization and can prevent the array from properly discriminating two close sound sources. This alteration of the localization is confirmed when considering *the same* two sources as above, emitting from a distance $r = 0.8m$ to the center of the array. The energy map in the farfield case turns into the one plotted in Figure 5. The low frequency components of the two sources are clearly in the nearfield of the array: the two main lobes of the energy map widen, caused by a weak directivity of the pattern, and deteriorates the quality of the localization. Moreover, the two sources are sometimes merged so that only one source is detected in a false direction. These simulations

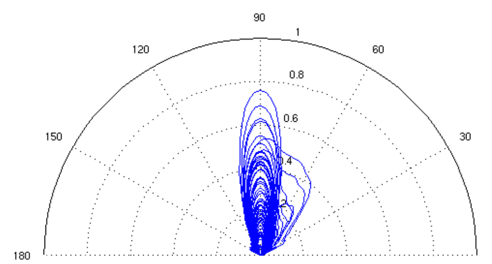


Fig. 5. Normalized energy map computed by a farfield beamformer over successive temporal windows (one curve per window) for nearfield sources.

show that all the improvements gained on the array synthesis in the farfield come to be wasted by using the same array in the nearfield. So, the idea outlined in this paper is to propose a new APS method which can be used to design an acoustic array in the nearfield. Basically, the method we described in [9] could also be used with Equation (3) to design a new APS method in the nearfield. But because of the high number of constraints involved into the optimization process, the mathematical conditioning of the method is not good and produces a lot of numerical problems. To solve these issues, we propose in the following a new optimization method, based on modal analysis, which is valid for farfield or nearfield beamforming.

III. Array modal analysis

A. Fundamentals

This section presents some elements of *Array Modal Analysis*, which unifies the formulation of the global response of a microphone array at all distances. In the following, a spherical coordinate system is considered where a point in space is represented by (r, θ, ϕ), with r and θ as in Figure 1, and ϕ the angle about the x-axis. The functions $P_m(.)$, $J_{m+\frac{1}{2}}(.)$ and $H^{(2)}_{m+\frac{1}{2}}(.)$ are respectively known as the *associated Legendre Function of the first kind*, the *half odd integer order Bessel Function of the first kind* and the *half odd integer order Hankel Function of the second kind*. The functions $h^{(2)}_m(.)$ and $Y_m(.,.)$ –respectively called the *spherical Hankel function of the second kind* and the *spherical harmonics* – are defined by

$$\begin{aligned} h^{(2)}_m(kr) &= \sqrt{\frac{\pi}{2kr}}\; H^{(2)}_{m+\frac{1}{2}}(kr) \qquad (4) \\ Y_m(\theta,\phi) &= \sqrt{\frac{2m+1}{4\pi}}\; P_m(cos\theta). \end{aligned}$$

By considering the general solution to the Helmholtz equation together with the normalizing function

$$\begin{aligned} R_m(r,k) &\triangleq re^{ikr}h^{(2)}_m(kr), \qquad (5) \\ \text{with } R_m(\infty,k) &\triangleq \lim_{r\to\infty} R_m(r,k) = \frac{i^{m+1}}{k}, \end{aligned}$$

[8] shows that nearfield and farfield beampatterns can be unified into a common formulation. When considering a linear array placed along the z-axis, any arbitrary beampattern $D(r,\theta,k)$, whether farfield or nearfield, can be rewritten as

$$D(r,\theta,k) = \sum_{m=0}^{\infty} \alpha_m(k) R_m(r,k) Y_m(\theta), \qquad (6)$$

where $\{\alpha_m(k)\}$ denotes a set of complex coefficients called *modal coefficients*. In fact, (6) defines an orthogonal transform pair analogous to the familiar Fourier series, in which

$$\alpha_m(k) = \frac{2\pi}{R_m(r,k)} \int_0^{\pi} D(r,\theta,k) Y_m(\theta) \sin\theta d\theta, \qquad (7)$$

so that the complex coefficients $\alpha_m(k)$ fully characterize $D(r,\theta,k)$.

B. Modal decomposition of a nearfield beamformer

The objective is to determine the expression of the modal coefficients of the general nearfield beamformer described by Equation (3), which assumes that each n^{th} filter $w_n(k)$ has the form (1). Recall that farfield equivalence is obtained by making $r \to \infty$. Equations (7) and (4) can be combined with Equation (3) to obtain, through the properties of the Legendre functions $P_m(\cos\theta)$, the final expressions of the modal coefficients for a linear array [8]

$$\alpha_m(k) = \gamma_m(k) \sum_{n=1}^{N} w_n(k) j_m(kz_n) \qquad (8)$$

where $j_m(x) \triangleq \sqrt{\frac{\pi}{2x}} J_{m+\frac{1}{2}}(x)$ are the *spherical Bessel* functions, and $\gamma_m(k) = -2ik\sqrt{\pi(2m+1)}$. Notice that the modal coefficients $\alpha_m(k)$ only depend on the array structure, i.e. on the sensor positions z_n and the type of filters $w_n(k)$. So, in the case when $w_n(k)$ is defined by (1), the modal coefficients $\alpha_m(k)$ can be written

$$\alpha_m(k) = \gamma_m(k)\; W^T J_m(k), \qquad (9)$$

$$\text{with } J_m(k) = V_{\text{Modal}}(m,k) \otimes V_{\text{FIR}}(k), \qquad (10)$$

$$\text{and } V_{\text{Modal}}(m,k) = \begin{pmatrix} j_m(kz_1) \\ j_m(kz_2) \\ \vdots \\ j_m(kz_N) \end{pmatrix}.$$

It is fundamental to understand that once the modal coefficients are known, the beampattern at any distance r can be computed through equation (6). Theoretically, (6) requires an infinite number of modal coefficients $\alpha_m(k)$. But a Parseval relation can be derived [8] to precisely characterize the error introduced by truncating this series up to a finite rank. In practice, few modal coefficients are sufficient to precisely approximate a beampattern, viz. there exists $M < +\infty$ such that

$$D(r,\theta,k) \approx \sum_{m=0}^{M} [\gamma_m(k) W^T J_m(k)] R_m(r,k) Y_m(\theta). \qquad (11)$$

For instance, by using Equation (11) with $M = 10$, the nearfield beampattern plotted through Equation (3) in Figure 4 can be approximated without any visible error. This property will be used in the following in order to get a more tractable optimization problem.

IV. Array Pattern Synthesis in the nearfield

It has been argued that any beampattern can be closely described by a finite set of modal coefficients $\alpha_m(k)$. The idea of the proposed method thus consists in minimizing the "worst-case distance" between the modal coefficients related to the unknown beampattern FIR filters through Equation (9), and these describing a reference beampattern $\widetilde{D}(r,\theta,k)$. The consequent optimization problem is presented in the next subsection. The modal decomposition of the reference beampattern and some optimization results follow.

A. Convex optimization problem

The aim of the proposed optimization problem is to determine the vector W made up with $N \times Q$ filters taps which enables to approximate the reference beampattern $\widetilde{D}(r,\theta,k)$ described by its complex modal coefficients $\widetilde{\alpha}_0(k), \ldots, \widetilde{\alpha}_M(k)$. This is dealt with through the following convex procedure:

$$\begin{aligned} &\text{minimize } \varepsilon \\ &\text{subject to } \|\; \alpha_m(k) - \widetilde{\alpha}_m(k) \;\| \le \varepsilon,\; \forall k \in K, \forall m \in \{0..M\}, \end{aligned} \qquad (12)$$

where K terms the set of selected frequencies over which the error minimization is performed. Although the number of coefficients required to *exactly* describe the obtained pattern $D(r,\theta,k)$ is infinite, the minimization is performed on a finite number $M+1$ of complex modal coefficients. Indeed, thanks to the Parseval relationship, the unconstrained modal coefficients $\alpha_m(k)$, with $m > M$, do not significantly affect the resulting array response. Finally, the constraint $W^T W < \delta$ is added to bound the array white noise gain from above [10], thus also limiting the admissible space of W.

B. Modal decomposition of the reference pattern

Every beampattern is periodic with respect to θ. So, it can be decomposed as a Fourier series on θ, included when the array is linear. To make the reference pattern $\widetilde{D}(r, \theta, k)$ at a distance r frequency-invariant, the following expression is proposed:

$$\widetilde{D}(r, \theta, k) = \text{Freq}(k) \sum_{l=0}^{\infty} a_l \cos(l\theta), \qquad (13)$$

where the scalar function Freq(k) depicts the array behavior along the temporal frequencies, and the remaining Fourier series is an even function of θ. In the following, to simplify notations, the case Freq$(k) = 1$ will be assumed. Noticeably, a pattern which is consistent with the small array size imposed by the robotics embeddability constraint, can be well-defined by truncating the Fourier series to a low number of terms. So, in the following, the basic idea of the proposed method is to keep only the first $L+1$ coefficients $\{a_0, \ldots, a_L\}$ in (13). By introducing this expression into (7), the modal coefficients of $\widetilde{D}(r, \theta, k)$ become

$$\widetilde{\alpha}_m(k) = \frac{\sqrt{\pi(2m+1)}}{R_m(r, k)} \sum_{l=0}^{L} a_l g_m(l), \qquad (14)$$

where $g_m(l)$, computed by induction on m, reads as:

$$g_m(l) = \frac{\big(l - (m-2)\big)\big(l + (m-2)\big)}{\big(l - (m+1)\big)\big(l + (m+1)\big)} \, g_{m-2}(l), \qquad (15)$$

$$\text{with} \begin{cases} g_0(l) = -\dfrac{1 + (-1)^l}{(l+1)(l-1)} \\ g_1(l) = \dfrac{-1 + (-1)^l}{(l+2)(l-2)}. \end{cases}$$

Further, $g_m(l)$ satisfies

$$\forall (m, l) \in \mathbb{Z}^2, \qquad \begin{cases} g_{2m}(2l+1) & = & 0 \\ g_{2m+1}(2l) & = & 0 \end{cases} \qquad (16)$$

$$\forall\ m > l, \qquad g_m(l) = 0. \qquad (17)$$

As an important consequence, if the Fourier series of the reference angular response is limited to its first $L+1$ terms, then the number of modal coefficients which are required to describe *exactly* such a pattern is also limited to $M + 1 = L + 1$.

C. Results

Consider the case when the objective is to listen to the direction $\theta_0 = 90°$ for frequencies $f \in [500 : 50 : 3000]$Hz. To compare with [7], the reference array response is parameterized by the angular bandwidth $\theta_p = 10°$ and the main lobe width $\theta_s = 26°$. The corresponding function of θ is then decomposed into its Fourier series, which is truncated to the order $L = 8$. The expressions of the coefficients $\{a_l, l \in [0, L]\}$, for $\theta_s \neq \theta_p$, are given by

$$\begin{aligned} a_0 &= \frac{\theta_p + \theta_s}{\pi}, \text{ and} \\ a_l &= \frac{4\cos(l\theta_0)(\cos(l\theta_p) - \cos(l\theta_s))}{\pi l^2 (\theta_s - \theta_p)}, \ \forall l > 0. \end{aligned} \qquad (18)$$

Then, the modal coefficients $\widetilde{\alpha}(k)$ are computed by using equation (14), with $r = 0.8m$. Because of the property (17), only $M+1 = L+1 = 9$ modal coefficients $\widetilde{\alpha}_0(k), \ldots, \widetilde{\alpha}_L(k)$ are sufficient to *exactly* obtain the reference beampattern. The filter orders are set to $Q = 150$. Once the modal coefficients $\widetilde{\alpha}(k)$ have been obtained, the optimization problem (12) is solved by means of the solver SDPT3 [11], coupled with YALMIP [12], under MATLAB. A vector W follows, which leads to the array response shown in Figure 6.

The resulting nearfield beampattern shows a constant main

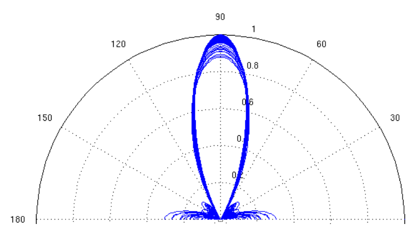


Fig. 6. Nearfield Optimized Filter-Sum Beamformer ($r = 0.8m$).

lobe width for all k over K, contrarily to the nearfield behavior of an array synthesized in the farfield, shown in Figure 4. Consequently, the resolution of the localization of a close speaker is expected to be better. This fact is confirmed when considering the localization process [7] which computes an acoustic energy map of the environment. For the same two sources as before, but emitting from the distance $r = 0.8m$, the energy map in Figure 5, obtained with a farfield beamformer, turns into the one plotted in Figure 7. In this case, the resolution of the localization has clearly increased and leads to an efficient detection of the two sources.

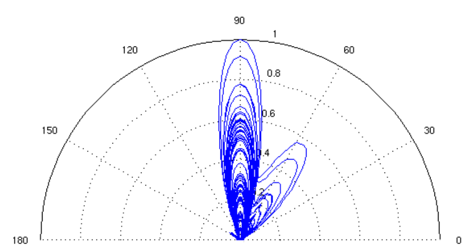


Fig. 7. Normalized energy map computed by a nearfield beamformer over successive temporal windows (one curve per window) for nearfield sources.

V. Discussion

In this section, the potentialities of the proposed optimization method are discussed. The approach is then compared to the farfield synthesis method taken from [7]. A discussion on nearfield properties and ongoing work follows.

A. Comparison of the APS methods

Though (12) and the farfield APS synthesis proposed in [9] both involve a frequency grid K, the use of modal coefficients leads to a new APS method which considers a continuum of angles θ. Consequently, the number of constraints decreases, so that the processing cost is lesser. Yet, in order to obtain quite similar results, the proposed APS method requires more FIR coefficients w_{nq} in the matrix W. Nevertheless, it seems less sensitive to the extra constraint $W^TW < \delta$, in that the optimization succeeds even with a low value of δ. With the same value of δ, the method [9] leads to a low quality array response which cannot be modified by increasing the number of FIR coefficients Q lest heavy numerical problems occur. As a consequence, the proposed approach appears to be more flexible (smaller constraint δ, and higher filter order Q) and more generic (valid formulation for farfield *or* nearfield operations).

B. Considerations on the nearfield case

Even though the sound signals of interest for speaker localization are mainly low frequency, beamforming techniques so far used in robotics rely on the farfield assumption. Figure 4 illustrates the inadequacy of such designs in the nearfield. Low frequencies, when emitted from a close distance, may appear in the nearfield, producing a poor focalization of the array. To evaluate the performance of a constant main lobe farfield beamformer as a function of the distance to the source, Figure 8 shows the distortion of the frequency component $500Hz$ of such a beampattern with respect to r. This alteration is continuous and such that the closer the speaker is, the higher is the sidelobe level. But for *this* frequency, if the speaker is far enough (approximately $r > 2m$), then the distortion induced by the nearfield can be neglected. However, one has

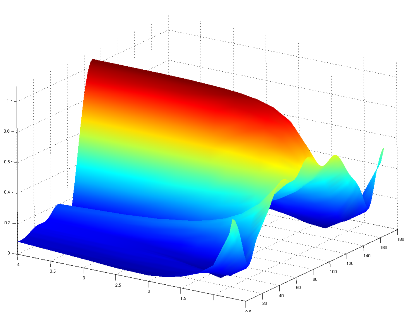


Fig. 8. $500Hz$ response of the farfield beamformer as a function of the distance r in meters.

to keep in mind that, despite the generality of the formulation, the optimization problem is explicitly solved for a given distance r. So, nearfield broadband beamforming can only be used to evaluate the angular localization of sound sources. Evaluating the range and bearing through beamforming *in real-time* is still an open problem. In the same way, designing a frequency-and-distance-invariant beampattern has no direct solution. Finally, a method allowing to filter both the direction and the distance has not been proposed so far.

C. Ongoing work

The experimental testbed is still under development. The $40cm$ long array is made up of 8 GRAS microphones. The acquisition card, which embeds all the required initial processing performs a synchronous analog-to-digital conversion with high resolution 18 bits delta-sigma data converters. The digital filters, designed through the optimization problem described so far, and the localization process, are implemented on a new generation VIRTEX FPGA equipped with DSP-cores for real-time processing. Low cost solutions based on DSPs have been tested, but this type of processor has proved to be inappropriate for massive parallel filtering. Finally, the resulting energy map is sent to the CPU embedded on the robot via USB.

VI. Conclusion

Using the theory of modal analysis, an original convex optimization procedure was designed to synthesize a broadband beampattern for speaker localization in the nearfield or in the farfield. The considered method allows to cope with specific requirements of mobile robotics such as real-time processing and small size of the microphone array. Simulation results have been proposed to illustrate the improvement of the approach with respect to preceding techniques, designed under the farfield assumption though used in the nearfield. The method presented here constitutes a reliable basis for developing higher level sound source localization and voice recognition algorithms for sensor-based control, or visuo-auditive tracking for human-robot interaction.

References

[1] A. A. Handzel and P. S. Krishnaprasad, "Biomimetic sound-source localization," *IEEE Sensors Journal*, vol. 2, pp. 607–616, 2002.

[2] J. Huang, K. Kume, A. Saji, M. Nishihashi, T. Watanabe, and W. Martens, "Robotic spatial sound localization and its 3D sound human interface," in *International Symposium on Cyber Worlds*, Nov. 2002, pp. 191–197.

[3] H.-D. Kim, J.-S. Choi, M. Kim, and C.-H. Lee, "Reliable detection of sound's direction for human robot interaction," in *IEEE/RSJ IROS*, vol. 3, Sept. 2004, pp. 2411–2416.

[4] D. Bechler, M. S. Schlosser, and K. Kroschel, "System for robust 3D speaker tracking using microphone array measurements," in *IEEE/RSJ IROS*, vol. 3, Sept. 2004, pp. 2117–2122.

[5] J.-M. Valin, J. Rouat, and F. Michaud, "Enhanced robot audition based on microphone array source separation with post-filter," in *IEEE/RSJ IROS*, vol. 3, Sept. 2004, pp. 2123–2128.

[6] Y. Tamai, Y. Sazaki, S. Kagami, and H. Mizoguchi, "Three ring microphone array for 3D sound localization and separation for mobile robot audition," in *IEEE/RSJ IROS*, Aug. 2005, pp. 903–908.

[7] S. Argentieri, P. Danès, P. Souères, and P. Lacroix, "An experimental testbed for sound source localization with mobile robots using optimized wideband beamformers," in *IEEE/RSJ IROS*, Aug. 2005.

[8] R. A. Kennedy, T. D. Abhayapala, and D. B. Ward, "Broadband nearfield beamforming using a radial beampattern transformation," *IEEE Transactions on Signal Processing*, vol. 46, no. 8, pp. 2147–2156, August 1998.

[9] S. Argentieri, P. Danès, and P. Souères, "Prototyping filter-sum beamformers for sound source localization in mobile robotics," in *IEEE ICRA*, 2005.

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

[11] R. Tutuncu, K. Toh, and M. Todd, *SDPT3 - A MATLAB Software Package for Semidefinite-Quadratic-Linear Programming*, http://www.math.nus.edu.sg/~mattohkc/sdpt3.html.

[12] J. Löfberg, *YALMIP : A Toolbox for Modeling and Optimization in MATLAB*, Taipei, Taiwan, 2004, available from http://control.ee.ethz.ch/~joloef/yalmip.php.