CONVEX OPTIMIZATION AND MODAL ANALYSIS FOR BEAMFORMING IN ROBOTICS: THEORETICAL AND IMPLEMENTATION ISSUES

Sylvain Argentieri† *and Patrick Danès*†

{*sargenti,danes*}*@laas.fr*

†*LAAS - CNRS, Université de Toulouse*
7, avenue du Colonel Roche, 31077 Toulouse, France

†*Université Paul Sabatier*
118, route de Narbonne, 31062 Toulouse, France

ABSTRACT

This paper deals with sound source localization in mobile robotics. This context opens new areas of research as it involves some specific constraints such as real time performance or embeddability. Quite often, broadband beamforming techniques are sought for, exploiting small-sized microphone arrays. A new approach to the design of broadband nearfield or farfield beamformers is first described, based on modal analysis and convex optimization. Original considerations related to the involvement of the theoretical beampattern into the real integrated acoustic sensor are then presented, which lead to alleviate some constraints during the optimization process and enable the use of smaller arrays. The whole method is illustrated on an example.

1. INTRODUCTION

For the last five years, the Robotics community has been concerned with the problem of sound source localization. The applications are many, e.g. exteroceptive robot control or visuo-auditive tracking, but they all require the availability of auditory cues at high rate. One classical way to tackle the problem consists in computing an acoustic energy map of the environment through a dedicated integrated sensor. Noticeably, in addition to temporal constraints, the robotics context raises other unexpected requirements. First, the sensor must be small enough to be embedded on a mobile platform; this fundamental need is seldom taken into account in usual signal processing approaches to the problem. Next, the aim is quite often to localize a human voice, an intrinsically broadband signal. Last, the relative distance between the sensor and a human speaker may cover a large range —from 50cm to 5m— so that low frequencies must be explicitly modeled as emitted in the nearfield.

Microphone array processing can lead to an efficient solution [1]. By steering broadband frequency-invariant beamformers to meaningful directions in the three-dimentionnal space, the above acoustic energy map can be readily computed. This strategy is commonly encountered in Robotics.

Convex optimization constitutes a generic framework to beamformer design. Such approaches have intensively been studied in the literature. Designing a microphone array under magnitude and phase response constraints has been cast into a convex problem in [2]. More specifically, spectral factorization has been used, in the case of equally spaced sensors, when no phase requirement is specified [3]. A convex formulation of the APS problem, under magnitude constraints only, was proposed in [2] for nonuniform arrays. More recently, an other semidefinite relaxation was proposed in [4] to solve a similar problem. Unfortunately, all these methods are illustrated in the narrowband case and cannot be directly used for the problem we consider. An application of convex optimization techniques to broadband array design was proposed in [5] by jointly optimizing the spatial and frequency response. A convex broadband APS method allowing to make the array sensitive to sound sources emitting only in the nearfield was proposed in [6]. The same problem was solved in [7] by means of linear programming techniques.

Many optimization-based methods minimize a distance between the actual beampattern and a given reference onto a finite grid in frequency and angular spaces. The array response over the unconstrained frequency domain must thus be carefully assessed, all the more if high pattern gains are likely to occur. Two approaches follow: on the one hand, structural constraints can be added right at the theoretical modeling stage so as to guarantee some properties at all frequencies [8]; on the other hand, the way the theoretical beampattern is involved in the effective response of the real integrated sensor can provide some guidelines to a less conservative statement of the synthesis problem, e.g. by relaxing constraints at "irrelevant" frequencies. This last idea can be benefited in the aforementioned Robotics context in order to reduce the size of the microphone array.

The paper is organized as follows. In section 2, a synthesis method is proposed to endow a linear array with a broadband frequency-invariant beampattern at a given distance r to the source —be this source in the nearfield or the farfield of the array. This method is based on an original combination of the modal analysis theory proposed in [9] with convex optimization arguments. Next, section 3 shows how the resulting theoretical filter-sum beamformer is entailed in the model of the real integrated sensor. Some guidelines to the selection of the optimization grid then follow, which are exploited in the case study being the subject of the section 4. A conclusion and some prospects end the paper.

2. A MODAL ANALYSIS BASED CONVEX OPTIMIZATION METHOD

In the following, a linear array of N sensors will be considered[1]. Each n^{th} sensor, placed at location z_n, is followed by

[1]Note that the proposed method can also be applied to multidimensional arrays by using their appropriate modal representation described in [9].

a digital Q^{th}-order FIR filter $BF_n(k)$. The N filters outputs are summed into the beamformer output. The response of such an array to a pointwise source operating at wavenumber k and located at polar coordinate (r,θ), with θ measured relative to endfire, is

$$D(r,\theta,k) = \sum_{n=1}^{N} BF_n(k) \frac{r\, e^{ik(r-d_n(r,\theta))}}{d_n(r,\theta)}, \tag{1}$$

where $d_n(r,\theta) = \sqrt{r^2+z_n^2-2rz_n\cos(\theta)}$ terms the distance from the source to the n^{th} sensor, and $i^2=-1$. By using modal analysis, such a beampattern can also be written [9]

$$D(r,\theta,k) = \sum_{m=0}^{\infty} \alpha_m(k) R_m(r,k) Y_m(\theta), \text{ with} \tag{2}$$

$$\alpha_m(k) = \frac{2\pi}{R_m(r,k)} \int_0^{\pi} D(r,\theta,k) Y_m(\theta) \sin\theta d\theta, \tag{3}$$

where $\{\alpha_m(k)\}$ denotes a set of modal coefficients. Equations (2) and (3) define an orthogonal transform pair which is analogous to the familiar Fourier series. The functions $Y_m(\theta)$ and $R_m(r,k)$ are defined by

$$\begin{aligned} Y_m(\theta) &= \sqrt{\frac{2m+1}{4\pi}}\, P_m(\cos\theta), \\ R_m(r,k) &= re^{ikr} h_m^{(2)}(kr), \end{aligned} \tag{4}$$

with $P_m(\cos\theta)$ and $h_m^{(2)}(kr)$ the Legendre and spherical Hankel functions, respectively. The modal coefficients of the linear array of interest can be easily shown to have the form

$$\alpha_m(k) = \gamma_m(k) \sum_{n=1}^{N} BF_n(k) j_m(kz_n), \tag{5}$$

where $\gamma_m(k) = -2ik\sqrt{\pi(2m+1)}$, and $j_m(kz_n)$ denotes the spherical Bessel function. Notice that they only depend on the array structure. By denoting $\otimes$ the Kronecker product, together with

$$\begin{aligned} W &= (w_{10}, w_{11}, \ldots, w_{1Q}, \ldots, \ldots, w_{N0}, \ldots, w_{NQ})^T, \\ V_{FIR}(k) &= (1, e^{-ikcT_s}, \ldots, e^{-ikcQT_s})^T, \\ V_{Modal}(k) &= (j_m(kz_1), j_m(kz_2), \ldots, j_m(kz_n))^T, \\ J_m(k) &= V_{Modal}(k) \otimes V_{FIR}(k), \end{aligned} \tag{6}$$

where w_{nq} is the q^{th} coefficient of the n^{th} sensor and T_s the sampling period, Equation (5) can be turned into the matrix form

$$\alpha_m(k) = \gamma_m(k)\, W^T J_m(k). \tag{7}$$

Theoretically, (2) requires an infinite number of modal coefficients $\alpha_m(k)$. But a Parseval relation can be derived [9] 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

$$D(r,\theta,k) \approx \sum_{m=0}^{M} [\gamma_m(k) W^T J_m(k)] R_m(r,k) Y_m(\theta). \tag{8}$$

Since all the properties of a given array pattern $D(r,\theta,k)$ are captured by its related modal coefficients, the idea of the proposed synthesis method is to minimize the distance between these actual modal coefficients and those describing a reference pattern $\widetilde{D}_r(\theta,k)$. Thanks to the 2π-periodicity and parity of any linear array beampattern w.r.t. θ, the following expression can be proposed for a reference beampattern at distance r :

$$\widetilde{D}_r(\theta,k) = \sum_{l=0}^{\infty} a_l(k) \cos(l\theta), \tag{9}$$

with $a_l(k)$ the coefficients of the Fourier series computed at frequency k. As the main features of any angular beamshape can be captured by a finite Fourier series, we assert in the following that a finite number of coefficients $a_l(k)$ are sufficient to define a meaningful reference beampattern. Consequently, only the first $L+1$ coefficients $a_0(k), \ldots, a_L(k)$ will be considered to describe $\widetilde{D}_r(\theta,k)$. By introducing Equation (9) into (2), the modal coefficients $\widetilde{\alpha}_m(k)$ of $\widetilde{D}_r(\theta,k)$ become

$$\widetilde{\alpha}_m(k) = \frac{\sqrt{\pi(2m+1)}}{R_m(r,k)} \sum_{l=0}^{L} a_l(k) g_m(l), \tag{10}$$

with

$$g_m(l) = \int_0^{\pi} \cos(l\theta) P_m(\cos\theta) \sin\theta d\theta. \tag{11}$$

Taking into account the Legendre functions properties, the following recursive equation satisfied by $g_m(l)$ in (11) has been computed:

$$g_m(l) = \frac{\big(l-(m-2)\big)\big(l+(m-2)\big)}{\big(l-(m+1)\big)\big(l+(m+1)\big)}\, g_{m-2}(l), \tag{12}$$

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

Further, $g_m(l)$ satisfies

$$\forall (m,l) \in \mathbb{Z}^2, \qquad \begin{cases} g_{2m}(2l+1) = 0 \\ g_{2m+1}(2l) = 0 \end{cases} \tag{13}$$

$$\forall\, m > l, \qquad g_m(l) = 0. \tag{14}$$

As an important consequence, if the Fourier series of the reference angular response is limited to its first $L+1$ terms, then the indices m of the modal coefficients required to describe exactly such a pattern are also limited to $m \in \{0;L\}$. To determine the vector W made of the $N\times(Q+1)$ FIR filters coefficients, the following convex optimization problem is proposed

$$\begin{aligned} &\text{minimize } \varepsilon \\ &\text{subject to } ||\, \alpha_m(k) - \widetilde{\alpha}_m(k)\, || \le \varepsilon, \\ &\qquad\qquad \forall k \in K, \forall m \in \{0,\ldots,M\}, \end{aligned} \tag{15}$$

where K terms the set of selected frequencies over which the error minimization is performed and $M \ge L$. (15) is a standard SOCP problem [10] which is convex and can then be efficiently worked out by specialized solvers such as SeDuMi[11] or SDPT3 [12], each of them being coupled with Yalmip [13] and Matlab. Finally, the constraint $W^TW \leq \delta$ is added to limit the admissible space of W, thus bounding the white noise gain.

3. PRACTICAL IMPLEMENTATION

In the sequel, $x_s(t) = \sum_p x[p]\delta(t - pT_s)$ terms the continuous-time counterpart of the digital signal $x[p]$ whose values are sampled from an analog signal $x(t)$ at period T_s. The Fourier transforms of $x(t)$ and $x_s(t)$ are respectively termed $X(f)$ and $X_s(f) = \frac{1}{T_s}\sum_m X(f - \frac{m}{T_s})$, with $f = kc/(2\pi)$ and $c = 340 m.s^{-1}$ the speed of sound. Figure 1 depicts a classical acquisition chain similar to the one embedded on our robotics platform. The source to be localized emits an analog broadband signal $e(t)$ which propagates to N microphones, whose identical frequency responses $M(f)$ reach up to 10kHz. Next, each microphone raw signal passes through an anti-aliasing filter $AA(f)$ before being sampled at T_s and processed by a digital filter with transfer function $G(z)$. Finally, the data are combined into the array digital output $s[p]$ through the N FIR filters $BF_1(z), \ldots, BF_N(z)$ determined by solving (15).

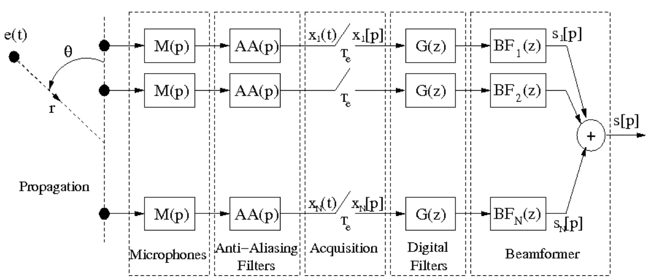

Figure 1: The data acquisition system

Though no transfer function can be defined between $e(t)$ and $s_s(t)$, the following relationships can be easily exhibited from Figure 1 for each n^{th} channel:

$$X_{n_s}(f) = \frac{1}{T_s}\sum_m ANA_n(f - \frac{m}{T_s})E(f - \frac{m}{T_s}), \quad (16)$$

$$S_{n_s}(f) = NUM_n(f)X_{n_s}(f), \quad (17)$$

with

$$ANA_n(f) = AA(f)M(f)\frac{r\, e^{i2\pi f(r - d_n(r,\theta))/c}}{d_n(r,\theta)}, \quad (18)$$

$$NUM_n(f) = BF_n(e^{i2\pi f T_s})G(e^{i2\pi f T_s}). \quad (19)$$

Noticing that $NUM_n(f) = NUM_n(f - \frac{m}{T_s})$, the above equations can be combined in order to express the Fourier transform $S_s(f)$ of the output signal $s_s(t)$ from the beamformer as

$$S_s(f) = \frac{1}{T_s}\sum_m S^\sharp(f - \frac{m}{T_s}), \quad (20)$$

with

$$S^\sharp(f) = \Big(\sum_{n=1}^{N} NUM_n(f)ANA_n(f)\Big)E(f) \quad (21)$$
$$= D(r,\theta,f)\big(AA(f)M(f)G(e^{i2\pi f T_s})E(f)\big).$$

Equation (20) shows that the spectrum $S_s(f)$, which ranges over constrained and unconstrained frequency sets, is the folding-and-adding of $D(r,\theta,f)E(f)$ beforehand filtered by the analog and digital filters of the acquisition chain.

It is worth considering these filters as extra degrees of freedom. Indeed, an explosion of the theoretical beamformer gain diagram on a subset K' of the complement of K defined in (15) can be tolerated as soon as they provide enough rejection over K'. For instance, assume that $AA(f)$ and $G(e^{i2\pi f T_s})$ show a flat gain response over $K = [500; 1000]$Hz and a strong rejection over K', with K' containing all frequencies above 1kHz, as indicated on figure 2. Despite the folding-and-adding (20), the behavior of the whole integrated sensor represented on figure 1 gets dicted by the beampattern on the domain K used for the optimization (15), even if this beampattern shows very high gains on K'.

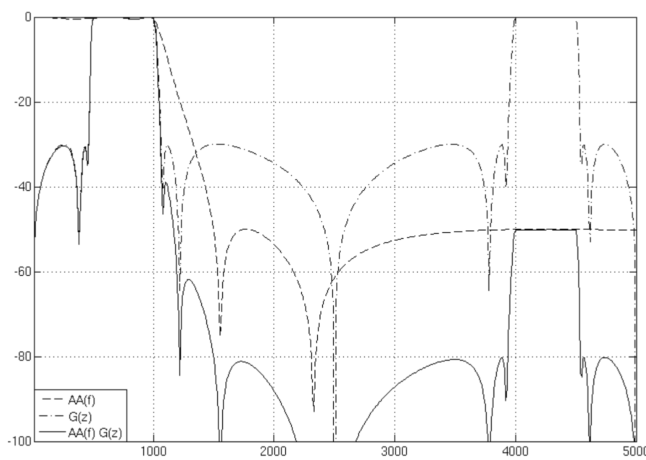

Figure 2: Typical frequency response of $AA(f)G(e^{i2\pi f T_s})$ for $M(f) = 1$ and $f_s = \frac{1}{T_s} = 5000$Hz: an explosion of the beampattern at frequencies higher than 1500Hz can be tolerated, for it implies no contribution to the frequency contents of $S_s(f)$ over $K = [500; 1000]$Hz due to the folding effect as soon as it is less than less than 50dB (see the response of $AA(f)G(e^{i2\pi f T_s})$ over $[f_s - 1000; f_s - 500]$Hz).

As an important consequence, relaxing some constraints on the array response on K' can enable a better fit with the reference on K, or a reduction of the spatial extent of the sensor. This fact is illustrated below.

4. CASE STUDY

In this section, a design example of a broadband nearfield frequency-invariant beampattern is considered. The first step is to define a reference beampattern $\widetilde{D}_r(\theta, k)$ which can be written

$$\widetilde{D}_r(\theta,k) = Freq(k)\sum_{l=0}^{L} a_l \cos(l\theta), \qquad (22)$$

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. To simplify, we will assume that $Freq(k) = 1$ for the selected frequencies over which the minimization (15) is performed. The proposed reference angular beamshape is rectangular, centered on $\theta_c = 90°$ and with a mainlobe width set to $2\theta_s = 30°$. Consequently, the Fourier coefficients a_l can be written

$$a_l = \begin{cases} \dfrac{2\theta_s}{\pi} & \text{for } l = 0 \\ \dfrac{2}{l\pi}\Big(\sin\big(l(\theta_c+\theta_s)\big) - \sin\big(l(\theta_c-\theta_s)\big)\Big) & \text{for } l > 0. \end{cases}$$

The Fourier series of $\widetilde{D}_r(\theta,k)$ is truncated up to the first $L+1 = 13$ terms. Notice that although the number of coefficients (7) required to exactly describe the pattern $D(r,\theta,k)$ is infinite, the minimization (15) is performed on a finite number $M+1$ of modal coefficients. Indeed, thanks to the Parseval relationship given in [9], the unconstrained modal coefficients $\alpha_m(k)$, with $m > M$, do not significantly affect the resulting array response by choosing an appropriate M. For this example, $M+1 = 19$ is used.

The next step consists in defining the frequency band of interest K. In practice, a frequency sampled version of the minimization problem (15) must be considered. Consequently, (15) is solved under Matlab with the solver SDPT3 [12] by considering 15 equally spaced frequencies in the set $K = [f_L; f_U]$, with $f_L = 500$Hz and $f_U = 1$kHz. Furthermore, the synthesis distance in this case study is set[2] to $3\lambda_L$.

Finally, the array parameters are chosen so that the order of the FIR filters $BF_n(k)$ and the sampling frequency f_s are respectively set to $Q = 32$ and $f_s = 5kHz$. A number $N = 15$ of sensors evenly-spaced at $\Delta z = \lambda_U/2$ are used[2] so that the length of the array is about 2.4 meters. Note that the proposed approach can also be used in the case of non-uniformly spaced sensors. The last parameter δ of the optimization problem (15) constraining the unknown weight vector W is set to $\delta = 250$.

The resulting beampattern, presented on Figure 3 for 25 frequencies in K, satisfies the requirements: it shows a frequency-invariant mainlobe, together with a sidelobe level smaller than about -26dB. Further, because the optimization procedure (15) minimizes the differences between two sets of modal coefficients only over a finite frequency set K, the beampattern explodes for high frequencies out of K, see figure 4. In practice, this phenomenon appears to be closely related to the constraint $W^T W < \delta$ joined to (15), in that the lower is δ, the more limited is the explosion. Nevertheless, this numerical explosion of the pattern can be thwarted by specifying minimal values for the stopband attenuation and transition band of the filters $AA(f)$ and $G(z)$. If the constraint $W^T W < \delta$ is suppressed, a greater stopband attenuation is needed to avoid the aliasing, which may lead to infeasible filters. Note that a short transition band is also required to counteract the slope of the numerical explosion at high frequencies (about 25dB over only 100Hz for the design example in Figure 4). In conclusion, the constraint $W^T W < \delta$, needed to limit the admissible space of W when solving (15), turns out to be also essential for a practical implementation of the proposed method. For example, when considering the technical data of the anti-aliasing filters used in the acquisition chain embedded on our robotic system —8$^{\text{th}}$-order lowpass ($f_c = 1$kHz) elliptic switched-capacitors filter from MAXIM with 50dB stopband rejection and a transition ratio of 1.5— the filters determined by solving (15) can be used to polarize the array when using an elliptic passband ([500; 1000]Hz) digital filter $G(z)$ giving at least 30dB of stopband attenuation for all frequencies greater than 1050Hz.

[2] λ_L and λ_U are the wavelengths respectively associated to the frequencies f_L and f_U.

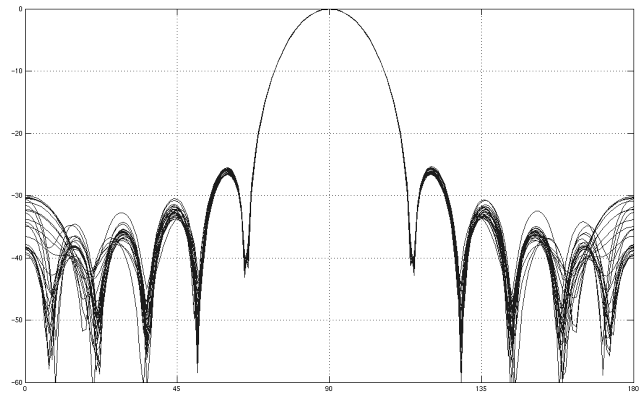

Figure 3: Normalized optimization result of a broadband nearfield frequency-invariant beampattern at radial distance $r = 3\lambda_L$ for 25 frequencies in $K = [500; 1000]$ Hz.

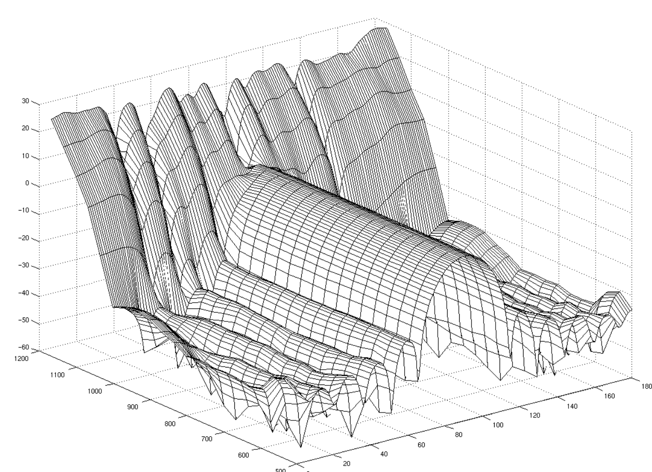

Figure 4: Numerical explosion of the obtained beamformer for unconstraint frequencies.

Performing the beamformer optimization in close connection with the study of its involvement in the real integrated sensor led to quite the same array response as [9], in terms of main lobe width and sidelobe level, for identical array parameters (i.e. FIR filter orders Q, frequency band K and sampling frequency f_s). But contrarily to [9] where a 29-sensors array of 3.4 meters is required, our approach benefits from the thwarted instability in K' to obtain a similar practical angular response with only 15 evenly spaced microphones, leading to a 2.4 meters long array. So we claim that this nice property is very useful when dealing with specific robotics constraints such as embeddability problems. Concerning the synthesis method itself, it is performed in one step whereas former nearfield optimization procedures based on modal analysis consist in a nearfield-to-farfield transformation followed by a farfield synthesis. In addition, compared with other convex optimization problems in the nearfield, the discretization of the angular space is not required. Thus, the number of constraints of the optimization problem (15) remains quite low, reducing the computational cost.

5. CONCLUSION

A new convex optimization method based on modal analysis has been proposed to design broadband nearfield or farfield filter-and-sum beamformers. A detailed description of the acquisition chain allowing the practical implementation of the method was given. We are currently working on the design of an acoustic array embedded on a robotic platform (see Figure 5). It is made of $N = 8$ omnidirectional microphones whose outputs are sampled and then processed by a FPGA-based board. The objective is to make the array sensitive to sound sources in the frequency band $[300; 3000]$Hz in order to localize human speakers. The method was successfully applied over this frequency domain for a 40cm long array which can be therefore easily embedded. Our objective is to perform in a close future the real-time processing required by classical robotics applications such as audio-visual tracking for human-robot interaction.

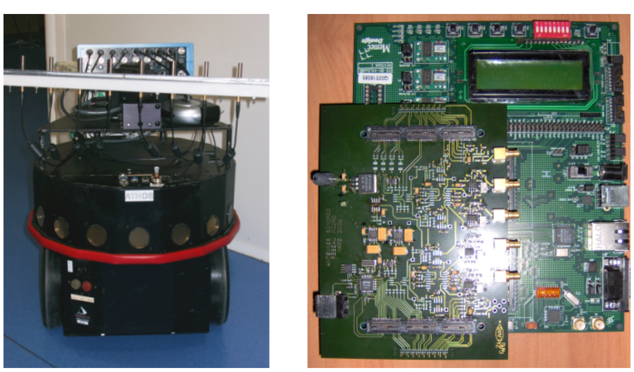

Figure 5: The Scout robotics platform equipped with a 8-microphones array (left). On-board acquisition system and FPGA processing card (right). Still in development.

REFERENCES

[1] B.D. Van Veen and K.M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, April 1988.

[2] Fan Wang, Venkataramanan Balakrishnan, Philip Yuanping Zhou, Jason Jiangnan Chen, Roy Yang, and Colin Franck, "Optimal array pattern synthesis using semidefinite programming," *IEEE Transactions on Signal Processing*, vol. 51, pp. 1172–1183, 2003.

[3] Shao-Po Wu, Stephen Boyd, and Lieven Vandenberghe, "Fir filter design via semidefinite programming and spectral factorization," in *Proceedings of Decision and Control*, Dec. 1996, vol. 1, pp. 271–276.

[4] Peter Kassakian, "Magnitude least-squares fitting via semidefinite programming with applications to beamforming and multidimensional filter design," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 2005, vol. 3, pp. 53–56.

[5] Dan P. Scholnik and Jeffrey O. Coleman, "Superdirectivity and SNR constraints in wideband array-pattern design," in *IEEE Radar Conference*, May 2001, pp. 191–186.

[6] Hervé Lebret and Stephen Boyd, "Antenna array pattern synthesis via convex optimization," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 526–531, March 1997.

[7] Ka Fai Cedric Yiu, Xiaoqi Yang, Sven Nordholm, and Kok Lay Teo, "Near-field broadband beamformer design via multidimensional semi-infinite-linear programming techniques," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 725–732, 2003.

[8] Darren B. Ward, Rodney A. Kennedy, and Robert C. Williamson, "Theory and design of broadband sensor arrays with frequency invariant farfield beam patterns," *Journal of the Acoustical Society of America*, vol. 97, pp. 1024–1034, Feb. 1995.

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

[10] Zhi-Quan Luo, *Mathematical Programming : Applications of convex optimization in signal processing and digital communication*, vol. 97, Springer-Verlag GmbH, July 2003.

[11] J. F. Sturm, "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11, pp. 625–653, 1999.

[12] R.H. Tutuncu, K.C. Toh, and M.J. Todd, *SDPT3-A MATLAB Software Package for Semidefinite-Quadratic-Linear Programming*, 2006, Available from http://www.math.nus.edu.sg/˜mattohkc/sdpt3.html.

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