An Experimental Testbed for Sound Source Localization with Mobile Robots using Optimized Wideband Beamformers

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

***Abstract*— This paper addresses the problem of practically implementing an original sound source localization strategy for mobile robots applications. The proposed method is based on a convex optimization solution to beamforming. It allows the sensing of signals within a direction of arrival and frequency domain of interest while rejecting other data. A precise description of the acquisition chain is proposed and a careful mathematical modeling is given in order to bridge the gap between theory and practical implementation. Simulation results and comparisons with classical filter-sum beamformer techniques are provided at the end of the paper to illustrate the performance of the sensor.**

***Index Terms*— Microphone arrays, beamforming, sound-source localization, mobile robotics.**

I. Introduction

In order to increase the ability of mobile robots to interact with humans, the need for developing performant auditory systems is preponderant. The constraint of processing real-time information relative to a changing and noisy environment makes the problem particularly challenging. For relevant applications, such as voice localization and processing for robot navigation in interaction with humans, the acoustic sensor must be able to localize in real-time sound sources within a wide frequency domain while rejecting other signals. In this sense, narrowband methods appear of limited interest. Considering that two ears enable most animals to perceive their sound environment accurately, numerous robotics auditory systems have been designed with only two microphones. To localize sound sources over a prescribed frequency band, such systems use the difference in phase (IPD) and in intensity level (IID) between the two sensors. The analytical determination of these interaural functions was proposed in [6] for a biomimetic artificial head modeled as a sphere. These difference functions are usually derived from the computation of the *head related transfer function* (HRTF) [4] [7], which requires precise measurements in an anechoic room. As such methods have proved to be highly sensitive to ambiant noise and changes in the environment, their application to mobile robotics turns out to be limited. Though the search for developing biologically inspired auditory systems for mobile robots, and especially for humanoids, is very motivating, it has been so far limited by strong technical issues. From an engineering point of view, the use of microphone arrays, including a larger number of sensors, still offers a more performant solution. Indeed, it allows to increase the resolution of the localization procedure and its robustness with respect to noise. Furthermore, sensor arrays design has been a long-time research topic in acoustics, and transferring a part of these techniques to mobile robotics is a promising way to offer high quality solutions in a near future. A robust sound source localization system, using a frequency-domain beamformer with a cubic array of eight microphones was developed in [13] for mobile robotics applications. On this base, geometric sound source separation and post-filtering techniques were proposed to isolate interfering speakers in a noisy environment [14]. Application of microphone arrays to talker localization and speech recognition was given in [5]. A 3D speaker tracking system using a multiple model adaptive estimator based on time delay of arrival was described in [2]. In this sense, we recently proposed a theoretical beamformer prototyping method for a linear microphone array, in the frame of our project for auditory system development [1]. The proposed approach is based on a convex optimization solution to beamforming and allows to make the antenna listen to a specific direction and select signals within a given wide frequency band. In this paper, we address the problem of technically implementing this audio processing method on

Fig. 1. Experimental Testbed. A 50cm array is being mounted on our SCOUT mobile robot

our robotics platform. A precise description of the acquisition chain is proposed and a careful mathematical modeling is given in order to brigde the gap between theory and practical implementation. The proposed beamforming method is recalled in § II. The modeling of the data acquisition system is described in § III. Finally, simulation results and comparisons with classical filter-sum beamformer techniques are given in § IV to illustrate the performance of the method.

II. A Broadband Array Pattern Synthesis Using Convex Optimization

a) Generalities: Under slight simplifying assumptions, the sound pressure field can be considered to be governed by a wave equation with a propagation speed of $c \approx 340\,\mathrm{m.s}^{-1}$ [3] [10]. One monochromatic solution of interest is the far-field solution with planar wavefront. In the following, f terms its temporal frequency and $\lambda = \frac{c}{f}$ its wavelength. No *a priori* statistical assumption is made on the informatory wave nor on the environment noise.

b) Array pattern: In the following, $.^T$ terms the non-Hermitian transpose operator. As shown in Figure 2, a rectilinear array of N passive omnidirectional nearly identical microphones is used, in which the i^{th} sensor is at location d_i on the x-axis of the world frame. All the microphones are assumed to have the same response $a(\theta, \phi, f) = a(f)$, with (θ, ϕ) the direction of arrival (DOA) of the incoming wave. Under the far-field assumption, the global response of the microphone array, or *array pattern*, is a function $P(\theta, \phi, f)$ and can be tuned up to a suitable design through *beamforming* [8] [15]. If the array elements are combined through gains $w_i(f)$, then it can be shown that the horizontal pattern $P_{\text{BEAMF}}(\theta, f) = P(\theta, \frac{\pi}{2}, f)$ is [10]

$$P_{\text{BEAMF}}(\theta, f) = \sum_{i=1}^{N} w_i(f) a(f) e^{j\frac{2\pi}{\lambda} d_i \sin\theta} = W(f)^T V(\theta, f), \quad (1)$$

where $V(\theta, f) = a(f)(e^{j\frac{2\pi}{\lambda} d_1 \sin\theta}, \ldots, e^{j\frac{2\pi}{\lambda} d_N \sin\theta})^T$ is the *steering vector* and $W(f) = (w_1(f), \ldots, w_N(f))^T$ is the gains column vector.

c) Filter-sum beamforming for broadband signals: To make the array sensitive to one spatial DOA θ_0 along the whole range of frequencies, one straight strategy is to compensate the time delays between sensors due to propagation by selecting the frequency-varying weights $W(f) = \frac{1}{N}(e^{-j\frac{2\pi f}{c} d_1 \sin\theta_0}, \ldots, e^{-j\frac{2\pi f}{c} d_N \sin\theta_0})^T$. The consequent beamformer —henceforth called "classical"— shows a main lobe uniformly centered on θ_0, whose width increases at low frequencies, see [1] for more details. More generally, beamforming techniques for broadband signals associate a separate K^{th}-order FIR to each of the N microphones. Consequently, the array pattern can be written as [11]

$$P_{\text{BEAMF}}(\theta, f) = \sum_{i=1}^{N} \sum_{k=1}^{K} a(f) w_{i,k} e^{-j2\pi f(k-1)T_e} e^{j2\pi f \tau_{\theta,i}}, \quad (2)$$

where $\tau_{\theta,i} = \frac{d_i}{c} \sin\theta$ terms the time delay from the emitting source to the i^{th} microphone due to wave propagation and $\{w_{i,k},\ k = 1, \ldots, K\}$ are the K^{th} taps of the FIR filter $BF_i(z)$ composing the i^{th} channel of the beamformer. Similarly to (1), the matrix formulation of this equation is[1]

$$P_{\text{BEAMF}}(\theta, f) = W^T V(\theta, f) \quad (3)$$

$$\text{with } V(\theta, f) = a(f)\ V_{Array}(\theta, f) \otimes V_{FIR}(f),$$

$$W = \begin{pmatrix} w_{1,1} \\ w_{1,2} \\ \vdots \\ w_{i,k} \\ \vdots \\ w_{N,K} \end{pmatrix},\ V_{FIR}(f) = \begin{pmatrix} 1 \\ e^{-j2\pi f T_e} \\ e^{-j2\pi f 2T_e} \\ \vdots \\ e^{-j2\pi f (K-1)T_e} \end{pmatrix}$$

$$\text{and } V_{Array(\theta,f)} = \begin{pmatrix} e^{j2\pi f \tau_{\theta,1}} \\ e^{j2\pi f \tau_{\theta,2}} \\ \vdots \\ e^{j2\pi f \tau_{\theta,N}} \end{pmatrix}.$$

d) A convex optimization framework to the design of beamformers: Denoting by $P_{\text{d}}(\theta, f)$ the desired array pattern, a synthesis strategy can be stated as the following optimization problem on the matrix W made of the $N \times K^{th}$ FIR taps:

$$\begin{array}{l} \text{minimize } \varepsilon \\ \text{subject to } |W^T V(\theta, f) - P_{\text{d}}(\theta, f)|^2 \leq \varepsilon,\ \forall (\theta, f) \in \Theta \times F, \end{array} \quad (4)$$

where Θ and F term the sets of DOA and frequencies over which the minimization is performed. This approach is inspired from [16], except that the temporal frequency dimension has been added. Because of the contraints it involves, (4) is a particular convex optimization problem, called second-order cone program (SOCP).

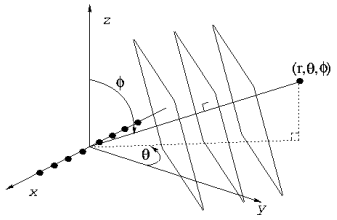


Fig. 2. Linear array in the far-field case

III. Modeling of the data acquisition system

This section describes the data acquisition system which is used for the development and the implementation of the speaker localization algorithm. The aim is to set up an exhaustive model taking into account the behavior of each element. The acquisition chain and the associated notations are represented in Figure 3.

[1] $\otimes$ represents the Kronecker product : given two matrices X and Y, whose respective dimensions are $p \times q$ and $t \times l$, the Kronecker product $X \otimes Y$ is the matrix Z of dimension $pt \times ql$ formed by the pq submatrices $Z_{i,j} = X_{i,j} Y$, for $(i, j) \in \{1, \ldots, p\} \times \{1, \ldots, q\}$.

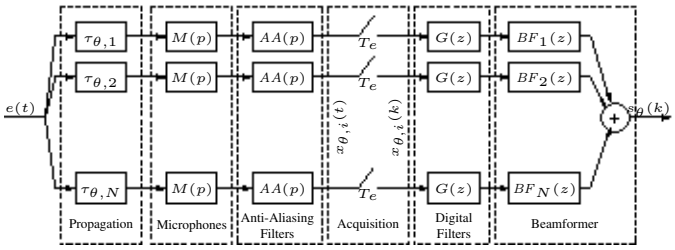


Fig. 3. The data acquisition system

A. Overview of the experimental setup

1) The incoming signal: The human voice is an acoustic signal which covers a wide frequency domain ranging from 100Hz to 20kHz [3]. Although its exact content varies from one person to the next, frequencies between 300Hz and about 3kHz are often considered sufficient to catch its main features. This is roughly the approximation used in telephone systems.

2) The microphones: $N = 8$ "GRAS Type 40PQ" microphones are used to build up the array. They are evenly spaced at $\frac{1}{2}\lambda_{3kHz} = 5.66cm$. Their frequency responses reach up to 10kHz. They are selected so that they match to each other in phase over their bandwidth with a $\pm 1^\circ$ tolerance. Consequently, they are henceforth assumed identical, with a common transfer function $M(p)$.

3) The acquisition board: The selected 8-channels "EZ-Kit Lite" acquisition board from Analog Device is equipped with 16-bits analog-to-digital converters operating synchronously at a maximum sampling rate of 48kHz. § III shows the relevance of the choice of the sampling frequency $f_e = 24$kHz. The audio data are then processed by one AD21065L DSP. Considering these DSPs characteristics, the localization algorithm is expected to run at the minimum frequency of 15Hz.

4) The anti-aliasing filter: Each microphone raw signal must pass through an anti-aliasing filter before being sampled by a board converter. The cutoff frequency of this filter —which must be low enough w.r.t. the Nyquist frequency $f_e/2$— is fixed to 3kHz, as justified in the next section. An 8^{th}-order lowpass elliptic switched-capacitors filter from Maxim is placed on each channel, whose transfer function is denoted by $AA(p)$.

5) The beamformer: The data coming from the 8 samplers are combined so as to make the array sensitive to incoming waves having prescribed frequencies and DOA. This property is achieved by associating to each i^{th} microphone a K^{th}-order digital FIR filter with transfer function $BF_i(z)$. The coefficients of $BF_1(z), \ldots, BF_N(z)$ are determined through convex optimization, as outlined in § II.

6) The digital filters: The used beamformer synthesis method enables both spatial filtering and —to a certain extent— temporal frequencies rejection. However, if too strong constraints are put on the desired antenna pattern, the underlying optimization program (4) may get ill-conditioned. One way to avoid this potential problem is to commit to an additional filter a part of the temporal filtering. So, identical digital band-pass filters with [300Hz-3kHz] bandwidth are placed just before every input to the beamformer. They are implemented as digital IIR filters with transfer function $G(z)$, whose passband ripple and stopband attenuation are specified in § III-B.2. Potential instability problems due to implementation round-off errors are expected not to occur since the aforementioned DSP provides a floating-point unit.

7) The output: The beamformer output, denoted by $s_\theta(k)$, is a digital signal which captures the contents of the incoming wave along the steered —or listened— direction θ. Its power is expected to be maximum when θ and the DOA θ^0 of the informatory incoming signal are identical. This assumption will be shown to hold, enabling source localization by scanning a whole range of hypothesized θ's.

B. Mathematical model

Before proceeding to the beamformer implementation, a further insight must be gained into this formula. This section thus aims at interpreting the way how (2) is involved into the behavior of the whole acquisition system depicted in Figure 3. All the anti-aliasing filters $AA(p)$ and passband digital filters $G(z)$ are assumed identical, which is a fair hypothesis since $G(z)$ is digital and since switched-capacitors anti-aliasing filters also have perfectly similar gain and phase responses. In the sequel, $x^*(t)$ terms the continous-time counterpart of the digital signal $x(k)$ whose values are sampled from an analog signal $x(t)$, viz. $x^*(t)$ is equal to the modulation $x^*(t) = \text{Ш}_{T_e}(t)x(t)$ of $x(t)$ by the Dirac comb $\text{Ш}_{T_e}(t) = \sum_n \delta(t - nT_e)$. The Fourier transforms —henceforth called "spectra", though with a slight language misuse— of $x(t)$ and $x^*(t)$ are respectively termed $X(f)$ and $X^*(f)$, so that $X^*(f) = \frac{1}{T_e}\sum_m X(f - \frac{m}{T_e})$.

1) Overall input-output relationship: As the beamformer output $s_\theta(k)$ is digital while the input signal $e(t)$ to the acquisition chain is analog, no transfer function can be defined between $e(t)$ and $s^*_\theta(t)$. However, for each i^{th} channel, the following relationships can be exhibited from Figure 3:

$$X^*_{\theta,i}(f) = \frac{1}{T_e}\sum_m ANA_i(\theta, f - \frac{m}{T_e})E(f - \frac{m}{T_e}), \quad (5)$$

$$S^*_{\theta,i}(f) = NUM_i(f)X^*_{\theta,i}(f), \quad (6)$$

with

$$ANA_i(\theta, f) = AA(f)M(f)e^{j2\pi f\tau_{\theta,i}} \quad (7)$$

$$NUM_i(f) = BF_i(e^{j2\pi fT_e})G(e^{j2\pi fT_e}). \quad (8)$$

Equations (5) and (7) indicate that the spectrum of $x^*_{\theta,i}(t)$ is obtained by folding-and-adding the spectrum of the input signal $e(t)$ beforehand propagated, aquired by the microphone and filtered through the anti-aliasing filter.

This confirms that $x^*_{\theta,i}(t)$ cannot be the linear filtering of $e(t)$. Besides, Equations (6) and (8) are $\frac{1}{T_e}$-periodic, for they relate the spectra of sampled signals. Noticing that $NUM_i(f) = NUM_i(f - \frac{m}{T_e})$, the above equations can be combined in order to express the spectrum $S^*_\theta(f)$ of the output signal $s^*_\theta(t)$ from the beamformer as

$$S^*_\theta(f) = \frac{1}{T_e} \sum_m \sum_{i=1}^{N} NUM_i(f - \frac{m}{T_e}) ANA_i(\theta, f - \frac{m}{T_e}) E(f - \frac{m}{T_e}), \quad (9)$$

which is just the folding-and-adding of the spectrum $S^\sharp_\theta(f)$ defined by

$$S^\sharp_\theta(f) = P_{\text{CHAIN}}(\theta, f) E(f), \quad (10)$$

with

$$\begin{aligned} P_{\text{CHAIN}}(\theta, f) &= \sum_{i=1}^{N} NUM_i(f) ANA_i(\theta, f) \quad (11) \\ &= \sum_{i=1}^{N} BF_i(e^{j2\pi f T_e}) G(e^{j2\pi f T_e}) AA(f) M(f) e^{j2\pi f \tau_{\theta,i}}. \end{aligned}$$

Comparing (11) with (2) while taking into account the equality $BF_i(e^{j2\pi f T_e}) = \sum_{k=1}^{K} w_{i,k} e^{-j2\pi f (k-1) T_e}$ leads to the conclusion

$$P_{\text{CHAIN}}(\theta, f) = \frac{S^\sharp_\theta(f)}{E(f)} = P_{\text{BEAMF}}(\theta, f) \quad (12)$$

provided that $a(f)$ is set to

$$a(f) = H(f) = G(e^{j2\pi f T_e}) AA(f) M(f). \quad (13)$$

2) Interpretation: Let's first consider the case when $a(f) = 1$. The optimization problem (4) is solved on the set defined by $F \times \Theta$. So, the difference between the pattern $P_{\text{BEAMF}} = W^T V$ and the desired pattern P_d is only minimized on this grid, leaving unconstrained all frequencies and angles out of F. For instance, the optimization results show that the array gain gets high for all frequencies upper than $\max(F)$. However, as indicated in § III-B.1, the spectrum $S^*_\theta(f)$, which ranges over constrained and unconstrained sets, is the folding-and-adding of $P_{\text{BEAMF}}(\theta, f) E(f)$. As a consequence, the array becomes sensitive to unexpected frequencies because of the folding of the pattern which lies in the unconstrained domain.

By Equation (10), the true array response involves P_{CHAIN} which includes the frequency response $H(f)$ of the acquisition chain. The degrees of freedom (DOF) in $H(f)$ can then be used to thwart the pattern explosion on the unconstrained set, thus avoiding the folding effect. First, a lower threshold can be specified for the absolute value of the stopband attenuation of $AA(f)$ and $G(f)$. However, because of the $\frac{1}{T_e}$ periodicity of $G(f)$, the attenuation of $H(f)$ is lesser at frequencies around f_e —see figure 4—, so that spurious peak values are still likely to occur in P_{CHAIN}. Consequently the second DOF, which is the sampling frequency, is risen up to 24kHz. Finally, the explosion in the above pattern P_{BEAMF} can be delayed so as to make $H(f)$ reach its maximum attenuation while P_{BEAMF} is low. It suffices to enlarge the frequency optimization set F by a bandwidth over which P_{BEAMF} should be close to $P_d = 0$.

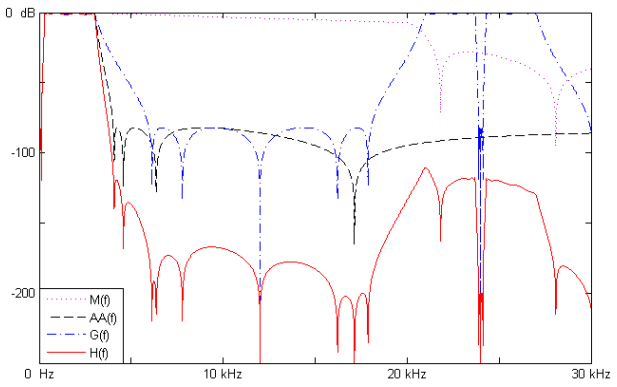


Fig. 4. Bode diagram of $M(f)$, $AA(f)$, $G(e^{j2\pi f T_e})$ and $H(f)$

As a consequence, the following parameters have been chosen for the filters (see figure 4 for details):

- $AA(f)$ is given by the Maxim datasheet, that is: 8^{th}-order lowpass elliptic filter with $0.4dB$ bandpass ripple, $82dB$ stopband rejection, the cutoff frequency $f_c = 3kHz$, and a sharp rolloff with a transition ratio of 1.5.
- $G(f)$ is a IIR bandpass digital elliptic filter, with $1dB$ bandpass ripple from $300Hz$ to $3kHz$, and $82dB$ stopband rejection.
- Microphones are modeled by a first order filter with a cutoff frequency of $10kHz$, followed by an element giving a $20dB$ rejection beyond $20kHz$.

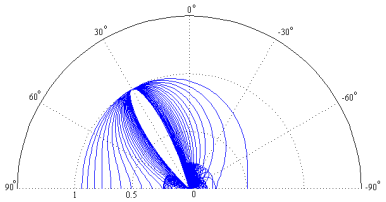
IV. Results

In this section, we first compare the patterns obtained through the optimization program (4) to classical beamformers. The spatial filtering efficiency of the localization procedure based upon such optimized patterns is then discussed.

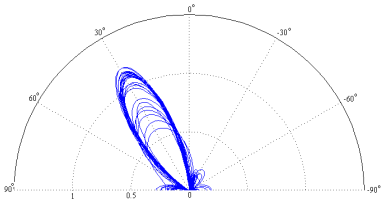
A. Optimized antenna patterns

Consider that the aim is to listen to the direction $\theta_c = 30°$ while keeping a constant main lobe width over the frequency interval $f \in [300Hz; 3kHz]$. The sets F and Θ in (4) are set to $F = \{300Hz, 400Hz, \ldots, 3kHz\}$ and $\Theta = \{-90°, -88°, \ldots, +90°\}$. In order to avoid temporal and spatial aliasing —see § III-B.2—, the desired pattern P_{d} is maintained at zero for each frequency f in $[3kHz; 3.5kHz]$. The angular bandwidth and the main lobe width of P_{d} —which are precisely defined in [1]— are respectively set to $\theta_p = 10°$ and $\theta_s = 24°$. The FIR filters order is $N = 50$. The problem (4) is solved using the solver SDPT3 (v3.2) [12] coupled with YALMIP [9], both of which run under MATLAB. The obtained pattern is drawn

in Figure 5 for each frequency $f \in F$. Figure 5(a) shows a pattern obtained through classical filter-sum beamforming. Its main lobe widens at low frequencies, for the array length is small compared to the corresponding high wavelengths. On the contrary, Figure 5(b) shows that the main lobe width of the optimized pattern remains nearly constant along the whole range of frequencies. This endows the array with a better resolution at low frequencies despite its interspacing is selected so as to respect the spatial Shannon sampling theorem, which is a contraint coming from high frequencies. Though oscillations in the main lobe level occur, inducing an attenuation at low frequencies, spatial filtering remains efficient.



(a) Classical Filter-Sum Beamformer

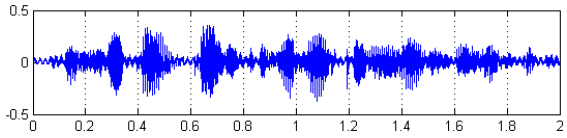


(b) Optimized Filter-Sum Beamformer

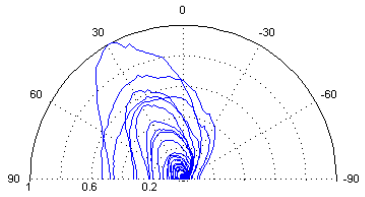
Fig. 5. Antenna pattern for f $\in F$

B. Localization results

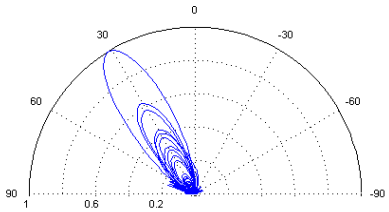
Once the optimization programs are solved for every θ, a bank of FIR filters is ready to be used so as to steer the array in any listened direction. A direct method for acoustic source localization immediately follows. It consists in making successive hypotheses on the DOA by scanning every θ in Θ. For each assumption, the beamformer's output energy is computed over a sliding temporal window with fixed width T_{obs}. In the sequel, we consider $T_{\text{obs}} = 1000T_e$. This leads to an *energy map*, which is expected to be maximum at the actual source DOA. In order to evaluate *in simulation* the performances of the optimized beamformer, we recorded on a basic microphone a male voice signal, presented in Figure 6(a). This evaluation is needful concerning localization and spatial focalization aspects. As the classical beamformer requires time delays which are not integer multiples of the sampling period, a digital approximation is implemented using FIR filters. The consequent pattern is similar to Figure 5(a), though at the expense of some spurious oscillations which somewhat contort the energy map in Figure 6(b).



(a) Voice signal



(b) Classical Filter-Sum Beamformer



(c) Optimized Filter-Sum Beamformer

Fig. 6. Normalized energy maps computed with classical and optimized beamformers over successive sliding temporal windows (one curve per window)

Figure 6 shows the energy maps computed over successive temporal windows. Consider first the classical filter-sum results in Figure 6(b). As human voice power is mainly located at low frequencies, the classical filter-sum pattern operates only with its low frequency components, producing a weak directivity which deteriorates the quality of speaker localization. Furthermore, this prevents the array to properly localize two persons speaking close to each other. So, classical beamforming does not seem to be appropriate for speaker localization, except in the case of large arrays, involving numerous microphones. On the other hand, Figure 6(c) shows that the optimized pattern leads to an efficient localization, producing energy maps with a nearly constant main lobe.

C. Spatial efficiency

To illustrate the spatial filtering efficiency of the above localization procedure, we now simulate two sound sources emitting from two distinct directions. The aim is to focus the array on the informatory source while rejecting the data coming from another DOA. In the sequel, we consider that the source of interest emits two sinusoids with frequencies $300Hz$ and $3kHz$, represented on Figure 7-1, from the DOA $\theta_{inf} = 30°$. A second source, considered as noise, emits $350Hz$, $550Hz$, $850Hz$, $4kHz$ and $5kHz$ sinusoids from the DOA $\theta_{noise} = -30°$. The microphones perceive

the sum of the multichromatic waves coming from the two emitting sources. The signal on microphone 1, represented on Figure 7-2, shows that the incoming information with 30° DOA is masked by the noise source.

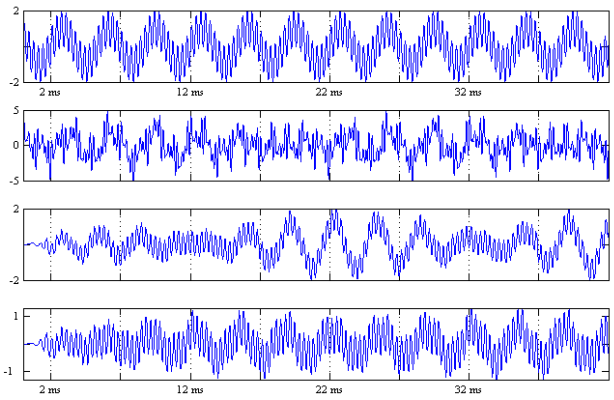


Fig. 7. Illustration of the spatial filtering efficiency on temporal plots

Figure 7-3 shows the output signal from a classical filter-sum beamformer steered to the DOA of the informatory source. The information captured in $e(t)$ is present, but modulated by low-frequency signals coming from the secondary source. In this case, the spatial filtering proves to be unefficient, as is corroborated by the energy maps on Figure 8 for the classical and optimized filter-sum beamformers. Clearly, the high width of the main lobe of the energy map corresponding to the classical beamformer leads to a poor quality focalization in the listened direction, which is the origin of the modulation seen on the temporal plots. On the opposite, a high spatial selectivity can be observed for the energy maps issued from the localization with the optimized filter-sum beamformer. The result shown in Figure 7-4 follows, in which a signal roughly similar to the informatory source can be recognized despite the presence of a *more powerful* noise. A transient state can be noticed lasting about $5ms$.

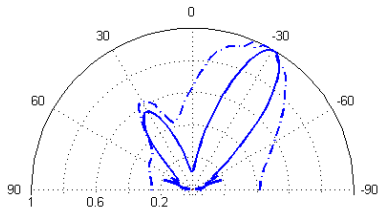


Fig. 8. Normalized energy maps for classical (dotted) and optimized beamformers

V. Conclusion

In this paper, we tried to give a precise description of the data acquisition chain which constitutes the implementation scheme of our speaker localization algorithm. The reasoning involved a careful mathematical modeling of the successive elements, which allowed to draw a parallel between the theoretical method and the practical development. Our experimental testbed will be completly operational in a very near future. The next step will be to complete our analysis with a large set of experimental tests. Then, the near-field case will be considered so as to deal with very close speakers. On this base, it will be possible to specify some important unknown parameters such as the minimal energy threshold which fires up the detection of a speaker in the environment of the robot.

VI. Acknowledgment

The work described in this paper was partially conducted within the project EGOCENTRE supported by the Interdisciplinary National Program in Robotics ROBEA of CNRS, and the EU Integrated Project COGNIRON ("The Cognitive Companion"), funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020.

References

[1] S. Argentieri, P. Danès, and P. Souères. Prototyping filter-sum beamformers for sound source localization in mobile robotics. In *IEEE International Conference on Robotics and Automation*, 2005.

[2] D. Bechler, M.S. Schlosser, and K. Kroschel. System for robust 3D speaker tracking using microphone array measurements. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2004.

[3] D. Brandstein, M. Wards, editor. *Microphone Arrays*. Springer, 2001.

[4] R. Brooks, T.B. Senior, and P.L.E. Uslenghi. The Cog project: Building a humanoid robot. In C.L. Nehaniv, editor, *Computations for Metaphors, Analogy, and Agents*. Springer Verlag, 1999.

[5] D. Giuliani, M. Omologo, and P. Svaiser. Talker recognition using a microphone array and a cross-powerspectrum phase analysis. In *International Conference on Spoken Language Processing*, 1994.

[6] A.A. Handzel and P.S. Krishnaprasad. Biomimetic sound-source localization. *IEEE Sensors Journal*, 2(6), December 2002.

[7] R.E. Irie. Multimodal sensory integration for localization in a humanoid robot. In *2nd IJCAI Workshop on Computational Auditory Scene Analysis*, 1997.

[8] H. Krim and M. Viberg. Two decades of array signal processing research. *IEEE Signal Processing Magazine*, July 1996.

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

[10] I.A. McCowan. *Robust Speech Recognition using Microphone Arrays*. Ph.D. thesis, Queensland University of Technology, 2001.

[11] D.P. Scholnik and J.O. Coleman. Formulating wideband array-pattern optimizations. In *IEEE International Symposium on Phased Array Systems and Technology*, 2000.

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

[13] J.M. Valin, F. Michaud, B. Hadjou, and J. Rouat. Localization of simultaneous moving sounds sources for mobile robot using a frequency-domain steered beamformer approach. In *IEEE International Conference on Robotics and Automation*, 2004.

[14] J.M. Valin, J. Rouat, and F. Michaud. Enhanced robot audition based on microphone array source separation with post-filter. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2004.

[15] B.D. Van Veen and K.M. Buckley. Beamforming : A versatile approach to spatial filtering. *IEEE ASSP Magazine*, April 1988.

[16] F. Wang, V. Balakrishnan, P. Zhou, J. Chen, R. Yang, and C. Frank. Optimal array pattern synthesis using semidefinite programming. *IEEE Transactions on Signal Processing*, 51(5):1172–1183, 2003.