

Conception d'un système de localisation de source sonore large bande par formation de voie pour des applications robotiques

Sylvain ARGENTIERI¹, Patrick DANÈS^{1,2}, Philippe SOUÈRES¹

¹LAAS - CNRS

7, avenue du Colonel Roche, 31077 Toulouse, France

²Université Paul Sabatier

118, route de Narbonne, 31062 Toulouse, France

Sylvain.Argentieri@laas.fr, Patrick.Danes@laas.fr, Philippe.Soueres@laas.fr

Résumé – Le travail présenté dans cet article a été mené dans le cadre d'un projet visant au développement d'un capteur auditif pour la localisation de locuteur en robotique mobile. La méthode de localisation proposée s'appuie sur une formation de voie large bande déterminée par résolution d'un problème d'optimisation convexe. Une description précise de la procédure d'optimisation et de la chaîne d'acquisition utilisée sur la plateforme expérimentale est donnée de façon à faire le lien entre la théorie et son implémentation pratique. Des résultats de simulation ainsi que des comparaisons avec les techniques classiques de formation de voie sont également fournis.

Abstract – The work presented in this paper comes as a part of a project which aims at developing an auditory system for a mobile robot in order to localize a speaker. The proposed localization method is based on a convex optimization solution to wideband beamforming. A precise description of the optimization procedure together with the acquisition chain of the experimental testbed is given in order to bridge the gap between theory and practical implementation. Simulation results and comparisons with classical beamforming techniques are also provided.

1 Introduction

Le groupe Robotique et Intelligence Artificielle du LAAS-CNRS est impliqué depuis de nombreuses années dans le développement de stratégies de navigation de robots mobiles. La plupart d'entre elles utilisent des modalités visuelles permettant de réaliser des tâches de localisation ou d'asservissement par rapport à des amers de la scène. Ces modalités sont ensuite complétées par l'ajout d'autres informations extéroceptives, e.g. fournies par des capteurs laser, ultrasoniques ou GPS [4]. Afin de compléter les capacités sensorielles de nos robots, nous travaillons à la conception et l'implémentation d'un système auditif embarqué.

Partant du constat que deux oreilles permettent à la majorité des animaux de percevoir efficacement leur environnement sonore, de nombreux systèmes de localisation de source sonore embarqués sur des plateformes robotiques exploitent seulement deux microphones [9]. Ces approches, dites "biomimétiques", s'appuient le plus souvent sur des mesures des différences en phase et amplitude des deux signaux reçus. Cependant, leur sensibilité importante aux variations des conditions expérimentales et au bruit ambiant limite leur intérêt dans un contexte robotique. Par conséquent, même si le développement de capteurs biomimétiques reste encore aujourd'hui un sujet de recherche très actif, il semble plus approprié, d'un point de vue ingénierie, d'utiliser une antenne de microphones [12]. Cependant, même si la théorie des antennes a déjà été largement étudiée sous des aspects Acoustique et Traitement du Signal [5], le contexte de la robotique mobile pose des problèmes inédits et originaux.

L'article est organisé comme suit. Nous détaillons d'abord

les spécificités propres au contexte de la robotique. Puis, dans une deuxième partie, nous présentons une méthode de synthèse d'antenne large bande basée sur la résolution d'un problème d'optimisation convexe. Enfin, dans une troisième partie, la chaîne d'acquisition utilisée sur la plateforme expérimentale est décrite, et des résultats de localisation en simulation sont apportés.

2 Spécificités liées à la robotique

En robotique, l'ensemble *capteur + système de traitement* est souvent un tout indissociable. On parle alors de *capteur intégré* dont l'exploitation matérielle et logicielle se trouve grandement simplifiée. Une architecture modulaire permettant d'intégrer facilement de nouvelles modalités sensorielles à nos robots est alors envisageable. Du fait que l'ensemble *antenne de microphones + algorithme de localisation* doit constituer un capteur sonore *autonome*, plusieurs impératifs techniques et théoriques sont à prendre en compte lors de sa conception.

Tout d'abord, l'aspect temporel apparaît comme un critère déterminant dans toutes les applications robotiques modernes. Ainsi, de nombreuses lois de commande sont basées sur l'introduction, dans la boucle d'asservissement, de données extéroceptives. Citons par exemple les méthodes d'asservissement visuel permettant d'asservir un robot sur un amer visuel [3]. Ceci est rendu possible par l'utilisation de capteurs délivrant l'information à une cadence suffisamment élevée pour réaliser une commande *temps réel*. L'introduction, dans une telle boucle d'asservissement, de données issues d'un capteur auditif n'est donc envisageable qu'avec un capteur *rapide*, capable

de fournir un résultat à une cadence minimale de 10Hz. Par conséquent, des méthodes de localisation de source très efficaces, telles que MUSIC ou Maximum de Vraisemblance [6], doivent être écartées. Au contraire, les méthodes de formation de voie, nécessitant moins de calculs en ligne, semblent mieux adaptées pour répondre à de fortes contraintes temporelles.

De plus, des contraintes d'embarquabilité fortes contraignent la dimension du capteur à synthétiser. En effet, compte tenu des dimensions des plateformes mobiles disponibles au LAAS, l'antenne de microphones ne pourra excéder une longueur de 50cm. De ce fait, les méthodes conventionnelles de formation de voie, qui compensent les retards causés par la propagation des signaux reçus sur chacun des microphones, possèdent une mauvaise résolution spatiale en basse fréquence et paraissent donc mal adaptées pour la localisation de signaux large bande tels que la voix.

Dès lors se pose la question de la conception d'une antenne acoustique de petite dimension utilisant une formation de voie large bande. Nous proposons dans la partie suivante une méthode permettant d'approcher cet objectif.

3 Synthèse d'une antenne large bande

3.1 Description du problème d'optimisation

Dans toute la suite, nous nous plaçons dans l'hypothèse de champ lointain et nous considérons une antenne linéaire de $N = 8$ microphones identiques omnidirectionnels, de fonction de transfert $a(f)$, placés à des abscisses d_i , $i \in \{1, \dots, N\}$. D'une manière générale, les méthodes de formation de voie pour signaux large bande combinent les sorties de filtres RIF d'ordre K placés en aval de chacun des microphones. Les coefficients de ces filtres étant calculés hors ligne, une implémentation sur DSP permet d'envisager une utilisation en temps réel du capteur acoustique. Soit θ l'angle d'incidence de l'onde considérée (voir figure 1) et f sa fréquence. L'objectif est de déterminer le vecteur $W = (w_{1,1}, \dots, w_{N,K})^T$ constitué des KN coefficients des filtres RIF, de façon que la réponse de l'antenne $P(\theta, f)$ dans un plan horizontal soit aussi proche que possible d'un comportement désiré $P_d(\theta, f)$. L'introduction des vecteurs $V_{RIF}(f) = (1, \dots, e^{-j2\pi f(K-1)T_e})^T$ et $V_{Antenne}(\theta, f) = a(f)(e^{\frac{2\pi f}{c}d_1 \sin \theta}, \dots, e^{\frac{2\pi f}{c}d_N \sin \theta})^T$ permet l'écriture suivante, où \otimes désigne le produit de Kronecker et $\tau_{\theta,i} = \frac{d_i}{c} \sin \theta$ [10] :

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

$$= W^T (V_{Antenne}(\theta, f) \otimes V_{RIF}(f)) = W^T V(\theta, f).$$

La procédure d'optimisation portant sur W , effectuée sur un produit $\Theta \times F$ d'ensembles discrets d'angles θ et de fréquences f est alors :

$$\begin{aligned} &\text{Minimiser } \varepsilon \\ &\text{Sous } |W^T V(\theta, f) - P_d(\theta, f)|^2 \leq \varepsilon, \forall (\theta, f) \in \Theta \times F. \end{aligned} \quad (2)$$

Une telle procédure d'optimisation a été récemment proposée dans [13] pour le cas faible bande et nous nous proposons

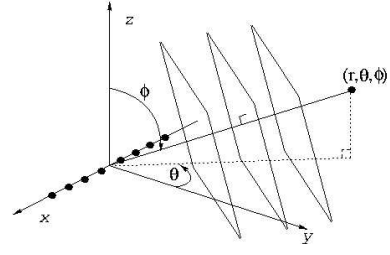


FIG. 1 – Antenne linéaire en champ lointain.

ici de l'étendre au cas large bande. Le comportement désiré $P_d(\theta, f)$ est supposé tenir compte des spécificités du contexte applicatif (petite taille de l'antenne, faible nombre de microphones), qui limitent de fait les performances pouvant être obtenues. Ce profil de référence est visible à la figure 2(a).

Ces contraintes se traduisent par la définition d'un lobe principal d'écoute dont la largeur est constante en fonction de la fréquence, ainsi que de zones devant être filtrées spatialement et fréquentiellement. Sa phase est quant à elle fixée à zéro. La largeur du lobe principal est paramétrée par les angles θ_c , θ_p et θ_s , représentant respectivement l'angle d'écoute, la bande passante et la bande coupée angulaire. Le problème ainsi formulé est un problème d'optimisation convexe [7] pour lequel il existe des solveurs efficaces capables de converger vers l'optimum global en temps polynomial.

3.2 Résultats d'optimisation

Considérons que notre objectif est d'écouter dans la direction $\theta_c = 30^\circ$, tout en conservant une largeur du lobe principal d'écoute constante sur la décade $f = [300Hz; 3kHz]$. Le problème d'optimisation (2) est posé sur l'ensemble $F \times \Theta = \{300Hz, 400Hz, \dots, 3.5kHz\} \times \{-90^\circ, -88^\circ, \dots, +90^\circ\}$. Le processus d'optimisation utilise le solveur SDPT3 (v3.2) [11] en association avec YALMIP [8] sous MATLAB. L'influence de chacun des degrés de liberté du problème d'optimisation a été précisément évaluée dans [1]. Nous avons montré que compte tenu du faible nombre de microphones utilisés, il est nécessaire d'effectuer un compromis entre l'efficacité du filtrage spatial (en terme de niveaux des lobes secondaire et largeur du lobe principal) et le gain de bruit d'antenne $B_a = W^T W$.

Un exemple de résultat d'optimisation, pour lequel $\theta_p = 10^\circ$, $\theta_s = 24^\circ$ et $K = 50$ est présenté figure 2(c), pour une antenne de $N = 8$ microphones régulièrement espacés de 5.66cm. Le lobe principal obtenu admet effectivement une direction et une largeur quasi-constantes en fonction de la fréquence, et ce même en basse fréquence malgré la petite dimension de l'antenne (45cm). Ainsi, par rapport à certaines techniques conventionnelles de formation de voie visant à compenser les retards temporels dus à la propagation, la résolution de la localisation d'un locuteur dont la voix comporte majoritairement des composantes basses fréquences pourra être améliorée, malgré la présence d'oscillations du niveau du lobe principal.

4 Modèle de la chaîne d'acquisition

Nous allons détailler dans cette partie la chaîne d'acquisition représentée figure 3, ainsi que le matériel de traitement

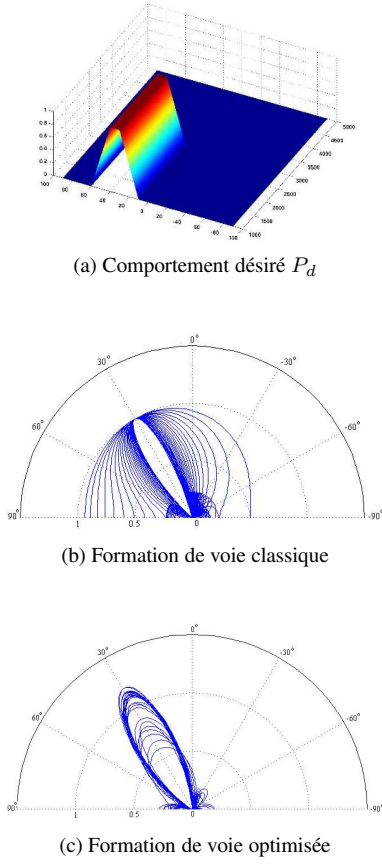


FIG. 2 – Réponses d’antenne pour $f \in [300Hz; 3kHz]$.

des données utilisé sur la plateforme expérimentale, de façon à comprendre comment la relation (1) est impliquée dans le comportement spatio-fréquentiel de l’antenne obtenue.

4.1 Matériel utilisé

Nous utilisons, afin d’acquérir les signaux émis par le locuteur à localiser, $N = 8$ microphones d’antennerie fabriqués par GRAS, appariés en phase à $\pm 1^\circ$, régulièrement espacés de $\frac{1}{2}\lambda_{3kHz} = 5.66cm$, et de fonction de transfert $M(p)$. Les signaux issus de ces microphones sont ensuite traités par un filtre anti-repliement à capacités commutées, de fréquence de coupure égale à $3kHz$ et de fonction de transfert $AR(p)$. L’utilisation de ce type de filtre permet de garantir que le filtrage effectué sur chacune des voies est totalement identique. Une fois les signaux filtrés, ceux-ci sont numérisés sur 16 bits par une carte d’acquisition d’Analog Devices équipée d’un DSP 21065L sur laquelle est implémentée la formation de voie. Celle-ci est réalisée par les filtres RIF de fonction de transfert $FV_i(z)$ déterminés hors ligne par la résolution du problème d’optimisation (2), mais aussi par des filtres RII de fonction de transfert $G(z)$ prenant en charge le filtrage permettant de ne retenir que la bande de fréquence d’intérêt (ici, la décade $[300Hz; 3kHz]$).

4.2 Description mathématique

Dans la suite, nous notons par $x^*(t)$ le signal temporel associé au signal numérique $x(k)$. Les transformées de Fourier, aussi appelées *spectres* par abus de langage, des signaux $x(t)$ et

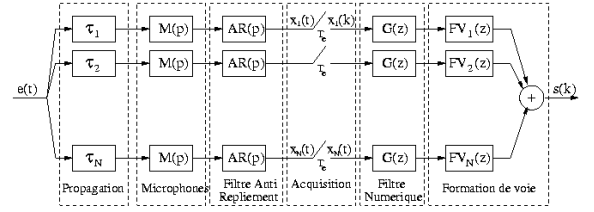


FIG. 3 – Chaîne d’acquisition

$x^*(t)$ sont respectivement notées $X(f)$ et $X^*(f)$, avec $X^*(f) = \frac{1}{T_e} \sum_m X(f - \frac{m}{T_e})$. Avec ces notations, on peut montrer que le spectre $S_\theta^*(f)$ du signal de sortie $s_\theta^*(t)$ s’exprime par :

$$S_\theta^*(f) = \frac{1}{T_e} \sum_m S_\theta^\#(f - \frac{m}{T_e}), \quad \text{avec} \quad (3)$$

$$S_\theta^\#(f) = P_{\text{CHACQ}}(\theta, f)E(f), \quad \text{et} \quad (4)$$

$$P_{\text{CHACQ}}(\theta, f) = \sum_{i=1}^N FV_i(e^{j2\pi f T_e})G(e^{j2\pi f T_e})AR(f)M(f)e^{j2\pi f \tau_{\theta,i}}$$

où $E(f)$ désigne le spectre du signal d’entrée. Comme cela était prévisible, il n’existe pas de fonction de transfert reliant le spectre continu du signal d’entrée au spectre numérique du signal de sortie. Pour autant, en comparant l’équation (3) avec (1) et sachant que $FV_i(e^{j2\pi f T_e}) = \sum_{k=1}^K w_{i,k} e^{-j2\pi f(k-1)T_e}$, nous avons :

$$P_{\text{CHACQ}}(\theta, f) = \frac{S_\theta^\#(f)}{E(f)} = P_{\text{FV}}(\theta, f) \quad (5)$$

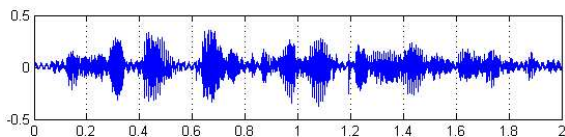
avec $a(f) = G(e^{j2\pi f T_e})AR(f)M(f)$. D’après (3), le résultat du problème d’optimisation (2) est donc replié et périodisé de façon à obtenir le spectre du signal de sortie $S_\theta^*(f)$. Il faut donc veiller à ce que ce repliement ne vienne pas modifier le comportement souhaité de l’antenne dans la bande des fréquences d’intérêt. Cela est rendu possible par un choix judicieux des caractéristiques des filtres $AR(f)$ et $G(z)$ qui vont permettre de contrer l’explosion numérique de la réponse d’antenne $P_{\text{FV}}(\theta, f)$ hors du domaine contraint. L’étude détaillée de la chaîne d’acquisition a été effectuée dans [2], et nous a permis d’aboutir aux paramètres suivants :

- $AR(f)$ est un filtre passe-bas elliptique du 8^{ème} ordre à capacités commutées possédant une fréquence de coupure $f_c = 3kHz$ et un niveau d’atténuation de $82dB$.
- La fréquence d’échantillonnage a été fixée à $f_e = 24kHz$.
- $G(z)$ est un filtre RII passe-bande elliptique, de bande passante comprise entre $300Hz$ et $3kHz$, apportant une atténuation de $82dB$ dans la bande coupée.

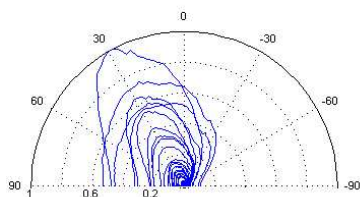
4.3 Résultats de localisation

Une fois la procédure d’optimisation effectuée pour chacune des directions d’écoute θ , nous disposons d’une banque de coefficients de filtres numériques capable de polariser l’antenne selon une direction souhaitée. Une méthode de localisation simple et peu coûteuse en temps de calcul consiste alors à écouter successivement dans toutes les directions plausibles d’arrivée de l’onde sonore. Pour chacune de ces directions, la puissance

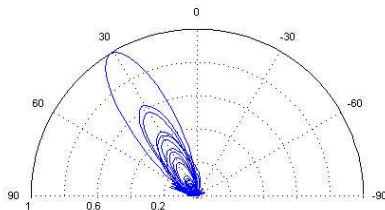
de sortie de la formation de voie est calculée sur une fenêtre temporelle notée T_{obs} , conduisant à la construction d'une carte d'énergie acoustique. Pour illustrer cette procédure de localisation, nous avons enregistré une voix masculine visible figure 4(a) sur un microphone de mauvaise qualité. La formation de voie classique est implémentée de manière numérique conduisant à une approximation de la réponse idéale 2(b) à l'origine des légères déformations constatées sur la figure 4(b).



(a) Signal vocal



(b) Formation de voie classique



(c) Formation de voie optimisée

FIG. 4 – Carte d'énergie normalisée, calculée pour des formations de voie classique et optimisée, sur une succession de fenêtres temporelles (une courbe par fenêtre).

Les figures 4(b)-(c) montrent les cartes d'énergie calculées successivement sur chacune des fenêtres d'observation de durée $T_{obs} = 1000T_e$. Considérons dans un premier temps le résultat de la localisation obtenu à partir d'une formation de voie classique, présenté figure 4(b). Comme la puissance de la voix est principalement localisée dans les basses fréquences, seuls des lobes possédant une largeur importante apparaissent dans le résultat de la localisation. En effet, seules les composantes basses fréquences de la réponse de l'antenne sont sollicitées, conduisant à une directivité médiocre à l'origine d'une détérioration de la qualité de la localisation du locuteur. La figure 4(c) montre quant à elle que l'antenne optimisée possède de fait une meilleure résolution de localisation, indiquée par une largeur de lobe principal quasi-constante, conséquence directe du comportement amélioré de la réponse de l'antenne en basses fréquences suite à la procédure d'optimisation.

5 Conclusion

Nous avons présenté une méthode de synthèse d'une formation de voie large bande par optimisation convexe adaptée à la localisation de locuteur. Malgré une antenne de petite dimension, les filtres numériques RIF obtenus permettent de localiser efficacement une source basse fréquence tout en garantissant une exécution en temps réel de la formation de voie. L'implémentation de la méthode sur la plateforme robotique expérimentale est en cours. Nous prévoyons en parallèle de revenir sur le conditionnement du problème d'optimisation proposé. De cette façon, nous espérons pouvoir tenir compte de l'incertitude sur la phase des microphones ainsi que réaliser des optimisations tenant compte uniquement des caractéristiques en gain du diagramme d'antenne souhaité.

Références

- [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] S. Argentieri, P. Danès, P. Souères, and P. Lacroix. An experimental testbed for sound source localization with mobile robots using optimized wideband beamformers. *IEEE International Conference on Intelligent Robots and Systems*, 2005.
- [3] D. Bellot and P. Danès. An LMI solution to visual-based localization as the dual of visual servoing. In *42nd IEEE Conference on Decision and Control*, pages 5420–5425, Maui, HI, 2003.
- [4] V. Cadenat, P. Souères, and M. Courdresses. Using system redundancy to perform a sensor-based navigation task amidst obstacles. *International Journal of Robotics and Automation*, 16(2), 2001.
- [5] J.C. Chen, K. Yao, and R.E. Hudson. Source localization and beamforming. *IEEE Signal Processing Magazine*, March 2002.
- [6] L. Kopp. *Détection et estimation en traitement d'antenne : applications*. Techniques de l'ingénieur - TE5226, 2003.
- [7] H. Leuret and S. Boyd. Antenna array pattern synthesis via convex optimization. *IEEE Transactions on Signal Processing*, 45(3) :526–532, March 1997.
- [8] J. Löfberg. *YALMIP : A Toolbox for Modeling and Optimization in MATLAB*, 2004. <http://control.ee.ethz.ch/~joloef/yalmip.php>.
- [9] K. Nakadai, D. Matsuura, H.G. Okuno, and H. Kitano. Applying scattering theory to robot audition system : Robust sound source localization and extraction. In *IEEE/RISJ International Conference on Intelligent Robots and Systems*, 2003.
- [10] D.P. Scholnik and J.O. Coleman. Formulating wideband array-pattern optimizations. In *IEEE International Symposium on Phased Array Systems and Technology*, 2000.
- [11] 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>.
- [12] J.M. Valin, F. Michaud, B. Hadjoui, 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.
- [13] 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.