
Cartographie d'un environnement sonore par un robot mobile

Auteur :

Aghilas SINI

Maîtres de stage :

Emmanuel VINCENT

François CHARPILLET

Rapport de stage de recherche au Loria, équipes PAROLE et MAIA

Master 2 en Intelligence artificielle, Reconnaissance des formes, Robotique

5 septembre 2014

Remerciements

Je tiens à remercier mes maître de stage Emmanuel Vincent et François Charpillet avec qui j'ai eu le privilège et l'honneur de travailler durant ce stage.

Je voudrais également remercier les membres de l'équipe MAIA (MAchines Intelligentes Autonomes) en particulier Nicholas Beaufort, Thomas Moinel, Maxime Rio, Jilles Di-bangoye, Xuan- Son Nguyen et Abdellah Dib qui m'ont beaucoup appris sur l'intelligence artificielle et la robotique.

Je remercie les membres de l'équipe Parole (Analyse, perception et la reconnaissance automatique de la parole) pour leurs accueil chaleureux.

Un grand merci à mes collègues de bureau Motaz Saad, Imran sheikh, Ilef Ben Ferhat, Arseiny Gorin, Jérémy Miranda et Guillaume Gris.

Mes remerciement vont également à Denis Jauvet, Antoine Luitkus, Nathan Souvira-a-Labastie, Luiza Orosanu, Yann Salaün, Dung Tran et Cyrine Nasri.

Je voudrai remercier Sarah Sini et Lynda Hadjeras pour leurs soutien.

Table des matières

Table des figures	iv
1 Formulation du problème	3
2 État de l'art	5
2.1 L'audition en robotique mobile	5
2.2 Localisation de sources sonores	6
2.2.1 Description de la méthode MUSIC	7
2.2.1.1 Rappel théorique de représentations des signaux et ma- trice de covariance	7
2.2.2 Méthode MUSIC	8
2.3 Modélisation de l'environnement de localisation	10
2.4 Conclusion	11
3 Approche Proposée	13
3.1 Grille d'occupation	13
3.2 Modèle de capteur	15
3.2.1 Description de l'expérience	16
3.2.2 Résultat de l'expérience	17
3.3 Mise à jour de la grille d'occupation	20
3.4 Commande du robot	20
3.4.1 Définition de la récompense	20
3.4.2 De la programmation dynamique pour la planification de trajectoire	22
3.4.3 Recherche du chemin de valeur optimale	25
4 Évaluation expérimentale	27
4.1 Protocole expérimental	27
4.2 Procédure expérimentale	30
4.3 Mesures d'évaluation	31
4.4 Résultat	32
5 Conclusion et perspectives	39
5.1 Conclusion	39
5.2 Perspectives	39

Bibliographie	41
----------------------	-----------

Table des figures

2.1	Réseau de microphones linéaire	6
3.1	Modèle de capteur $p[m_t d, \alpha]$ pour une source de direction d'arrivée $\alpha = 0^\circ$, pour différentes distances d.	17
3.2	Modèle de capteur $p[m_t d, \alpha]$ pour une source de direction d'arrivée $\alpha = 90^\circ$, pour différentes distances d.	17
3.3	Modèle de capteur $p[m_t d, \alpha]$ pour une source de direction d'arrivée $\alpha = 170^\circ$, pour différentes distances d.	18
3.4	Modèle de capteur $p[m_t d, \alpha]$ pour une source de direction d'arrivée $\alpha = 200^\circ$, pour différentes distances d.	18
3.5	Modèle de capteur $p[m_t d, \alpha]$ pour une source de direction d'arrivée $\alpha = 345^\circ$, pour différentes distances d.	19
3.6	Positions possibles du robot mobile p_{T+k} (pour une orientation donnée) à chaque instant $T + k$ dans le plan 2D	23
3.7	Graphe séquentiel associé au déplacement du robot mobile	24
4.1	Photo de l'appartement intelligent du centre de recherche INRIA de Nancy	27
4.2	Photo du robot expérimental TurtleBot 2	28
4.3	Capteurs du TurtleBot	28
4.4	Couverture de la grille par rapport à la pièce	29
4.5	Différentes configurations de la source sonore dans la grille d'occupation .	30
4.6	Modélisation de l'environnement du robot mobile	32
4.7	Résultats de la trajectoire optimale pour une source sonore à 1.20 m et 45°	33
4.8	Résultats de la trajectoire optimale pour une source sonore à 2.40 m et 45°	34
4.9	Résultats de la trajectoire optimale pour une source sonore à 1.20 m et 90°	35
4.10	Résultats de la trajectoire optimale pour une source sonore à 2.40 m et 90°	35
4.11	Résultats de la trajectoire optimale pour une source sonore à 2.40 m et 135°	36
4.12	Résultats de la trajectoire optimale pour une source sonore à 1.20 m et 135°	36

Introduction

Un robot mobile autonome comme agent intelligent face à des situations complexe doit avoir des capacités de perception, de pensée et d'action. Après de longues recherches, les propriétés individuelles du robot intelligent telles que la robustesse et la fiabilité ont été grandement améliorée.

Une des applications les plus importantes des robots mobiles est d'explorer un espace inconnu et d'exécuter une tâche dans un environnement dangereux.

La technologie de localisation de cible est l'un des aspects fondamentaux pour la navigation du robot mobile, utilisant des capteurs pour collecter des informations sur l'environnement et pour déterminer la position de la cible par les données perçues. L'un des capteurs les plus populaires utilisés aujourd'hui est le capteur visuel pour robot mobile. Le robot détecte généralement le monde extérieur avec des caméras, et il ne peut pas être en mesure d'utiliser le capteur visuel pour obtenir des informations fiables sur l'environnement lorsque l'objet n'existe pas dans le champ visuel de la caméra, ou lorsque les conditions d'éclairage sont faibles. De tels événements non visibles qui peuvent, toutefois, être accompagnée d'une émission sonore. Dans les situations ci-dessus les informations les plus utiles pour l'homme sont pour la plupart des données audio. Le système auditif humain est un système de traitement de l'information complexe à support biologique, il peut percevoir l'intensité du son et son orientation dans l'espace. Par rapport à la vision, l'audition a plusieurs propriétés uniques. L'audition est omnidirectionnelle. Quand une source sonore émet de l'énergie, le son remplit l'air, et un récepteur de son reçoit l'énergie sonore dans toutes les directions. Certains appareils spécialisés peuvent également recevoir une image de toutes les directions, mais encore il faudra tout de même balayer la surface totale pour localiser un objet spécifique. Ainsi, un robot peut percevoir l'information auditive d'une source derrière se trouvant derrière un obstacle. Même lorsque la source sonore est en dehors d'une chambre, le robot peut d'abord localiser la direction de la source de sonore, puis se déplacer et écouter de nouveau, jusqu'à ce que finalement il localise la source sonore. Par conséquent, la capacité d'audition d'un robot peut compenser les restrictions d'autres capteurs tels que la vue limitée ou les obstacles non

translucides. Ainsi, cette capacité augmente les types d'informations sur l'environnement acquises par le robot et améliore la précision de la prise de décision du robot. Par conséquent, la localisation de la source sonore est une technologie à très large éventail d'applications dans le domaine militaire et civil pour ses avantages uniques.

Chapitre 1

Formulation du problème

L'objectif de la localisation de sources sonores avec un robot mobile est d'exploiter la mobilité du robot pour localiser la position des sources sonores d'une manière précise. Un capteur équipé de plusieurs microphones capable d'enregistrer des sources sonores est placé sur le robot mobile. Les signaux enregistrés par ce capteur sont le résultat de la propagation et de la superposition des sources sonores. Le processus de localisation de sources sonores avec un robot mobile tient compte de plusieurs connaissances :

- des connaissances en robotique mobile,
- des connaissances sur le processus de localisation de sources sonores, le type et le modèle de capteur, la nature des sources sonores et la méthode de localisation.

L'autonomie de mouvement du robot mobile, c'est-à-dire la capacité d'effectuer un déplacement donné sans intervention extérieure, permet de localiser les sources sonores. La planification de trajectoire consiste à déterminer une trajectoire possible entre deux points en tenant compte des positions estimées des sources sonores. Le déplacement du robot mobile suppose donc au moins deux types de connaissances préalable :

- des connaissances sur l'environnement dans lequel évolue le robot mobile, éventuellement sur les sources sonores,
- des connaissances sur la position du robot mobile dans l'environnement.

Ces deux types de connaissances requièrent respectivement un processus de modélisation de l'environnement du robot mobile, et un processus de localisation du robot.

Dans ce rapport l'accent est mis sur la commande du robot mobile dans un environnement supervisé de type pièce expérimentale. Au sein de cet environnement, on peut distinguer entre d'une part, l'environnement statique —les murs de la pièce, qui sont parfaitement connus et la position des sources sonores présentes dans la pièce, dont on n'a aucune connaissance *a priori*—et, d'autre part, l'environnement dynamique—le robot mobile dont la position est supposée connue. De plus cet environnement est structuré, en ce sens qu'il est descriptible, en deux dimensions.

Chapitre 2

État de l'art

2.1 L'audition en robotique mobile

L'audition en robotique comme thème de recherche est assez récent. En comparaison avec les travaux réalisés en perception visuelle, l'étude de l'audition en robotique est encore à ses balbutiements. En 2004, La conférence IEEE IROS (*International Conference on Intelligent Robots and Systems*) a pour la première fois ouvert une session spéciale sur l'audition en robotique.

Les premiers travaux sur la localisation de sons remontent aux années 90 par Irie pour le robot Cog. Les capacités mises en œuvre étaient très limitées, en raison de limitations matérielles. Un peu plus tard dans les années 2000, le robot SIG et ses successeurs conçus à l'université de Kyoto ont considérablement augmenté leurs capacités auditives. Ces derniers sont construits sur la base de l'audition binaurale [1], qui est encore la forme la plus commune d'audition artificielle sur les robots mobiles. Les premiers travaux de Nakadai [2] sur l'audition active ont permis de localiser des sources sonores sur le plan horizontal à l'aide de l'audition binaurale. Depuis des travaux en mis l'accent sur la séparation de sources sonores [3][4] et la reconnaissance de la parole [2]. les robots commencent à tirer avantage de l'utilisation de plus deux microphones : c'est le cas par exemple de *Sony QRIO SDR4XII* avec sept microphones. Des logiciels dédiés à l'audition en robotique ont été développés récemment tel que le logiciel Hark, à l'Université de Kyoto. Ce logiciel libre permet l'analyse de la scène auditive en 3D, la localisation et le suivi de sources sonores en temps réel [5], la séparation de sources sonores et la reconnaissance de la parole.

2.2 Localisation de sources sonores

La localisation des sources sonores dépend de plusieurs facteurs : le nombre de microphones, la nature des sources qu'on souhaite localiser, le type de capteurs (microphone omnidirectionnel, Cardioïde ...), et leur qualité.

Dans [6] les méthodes de localisation ont été classées en trois catégories :

- les méthodes par histogramme,
- les méthodes par clustering,
- les méthodes par spectre angulaire, par exemple *MUSIC*, *GCC-PHAT*, *cSCT*.

L'objectif de ces méthodes est d'estimer la direction d'arrivée *DOA* (*Direction Of Arrival*) d'une ou plusieurs sources sonores sur une antenne de microphones en champ lointain.

Soit une antenne acoustique formée de n capteurs omnidirectionnels, identiques, ponctuels. Cette antenne est placée en champ lointain et captera donc des ondes planes. L'antenne et la source sont placées dans le même plan horizontal de sorte que la source sera repérée uniquement par l'angle polaire θ , angle que fait l'axe de l'antenne avec la direction du vecteur d'onde (2.1).

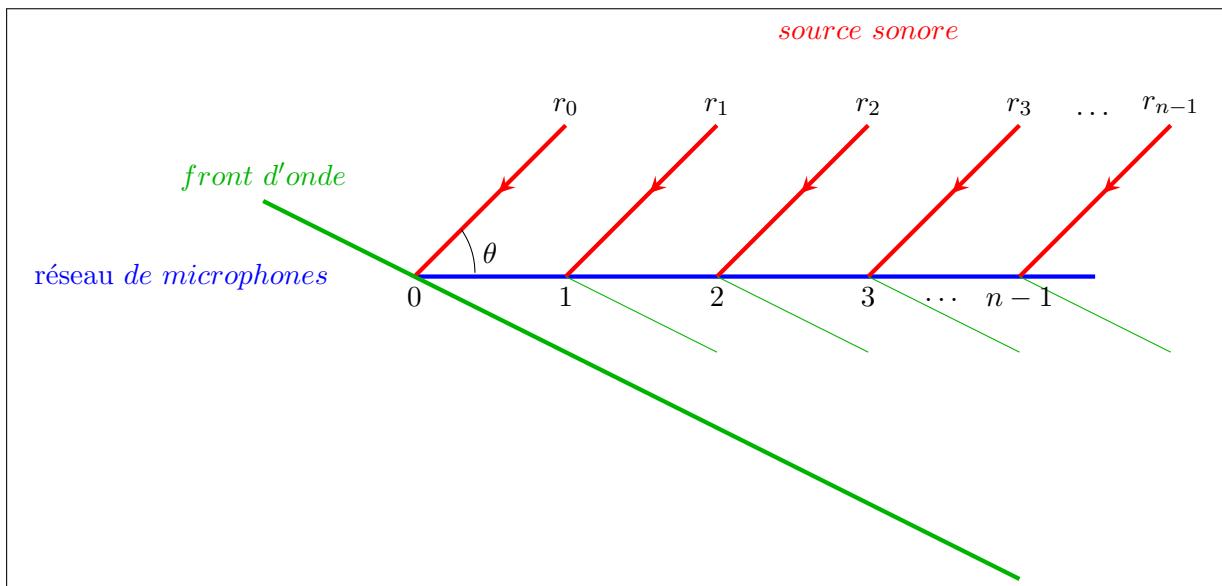


FIGURE 2.1: Réseau de microphones linéaire

L'une des méthodes les plus fréquemment utilisée pour la localisation de sources sonores en robotique mobile est la méthode *MUSIC* (*Multiple Signal Classification*).

Cette méthode s'appuie sur la décomposition spectrale de la matrice de covariance des signaux provenant des capteurs. Elle sépare l'espace caractéristique de la matrice de

covariance du signal global en sous-espace signal propre et sous-espace bruit en utilisant une décomposition en valeurs propres de la matrice de covariance. Cette méthode a été proposée par Schmidt [7] pour estimer les directions d'arrivée des signaux. Elle est capable de détecter plusieurs sources sonores et présente de meilleures performances que la méthode de formation de voies[8] conventionnelle à condition que ces sources soient dé-correlées ou faiblement corrélées. Si les sources sont fortement corrélées, cette méthode nécessite une modification par les algorithmes de dé-corrélation.

2.2.1 Description de la méthode MUSIC

La méthode *MUSIC* classique est basée sur l'analyse de la matrice de covariance spatiale des signaux issus des capteurs.

2.2.1.1 Rappel théorique de représentations des signaux et matrice de covariance

On considère une antenne composée de N capteurs omnidirectionnels. Cette antenne reçoit P ($P < N$) signaux émis par des sources rayonnantes auxquels se superpose un bruit additif. On suppose que les signaux émis par ces sources sonores sont stationnaires, centrés et non corrélés avec le bruit additif[9]. Les P sources sont placées dans un champ lointain, donc supposées ponctuelles, et le bruit additif observé sur les différents capteurs est stationnaire, centré, et de variance σ^2 . Les signaux reçus sur les capteurs de l'antenne peuvent être représentés par la combinaison linéaire [7] suivante :

$$x(t) = A(\theta)s(t) + b(t) \quad (2.1)$$

- $x(t)$ vecteur des signaux perçus par les capteurs de l'antenne,
- $s(t) = [s_1(t) \ s_2(t) \dots s_P(t)]^T$: vecteur des amplitudes complexes des signaux émis par les P sources à l'instant t ,
- $A = [a(\theta_1), a(\theta_2) \dots a(\theta_P)]$: matrice des vecteurs directionnels de dimension $(N \times P)$,
- $b(t) = [b_0(t) \ b_1(t) \dots b_{N-1}(t)]^T$: vecteur bruit additif sur les N capteurs.

Les P signaux des sources sont supposés indépendants. La matrice de covariance de ces sources est alors :

$$R_s = \mathbb{E}[s(t) \ s^H(t)] = diag(p_1, p_2 \dots p_P) \quad (2.2)$$

avec p_i puissance de la i ème source sonore que l'on souhaite localiser. On a supposé également que les signaux émis par ces sources sont stationnaires, centrés et non corrélés

avec le bruit (bruit blanc), on peut donc déduire la matrice de covariance des sorties des capteurs formant l'antenne :

$$R = E[x(t) \ x^H(t)] = A(\theta)R_sA^H(\theta) + R_b \quad (2.3)$$

où :

- $R_s = \mathbb{E}[s(t) \ s^H(t)]$: matrice de covariance des signaux des sources de dimension $P \times P$;
- $R_b = \mathbb{E}[b(t) \ b^H(t)] = \sigma^2 I$: matrice de covariance du bruit ambiant de dimension $N \times N$, où σ^2 est la variance du bruit et I est la matrice identité ;

En pratique, on estime la matrice de covariance à partir d'un nombre fini d'échantillons temporels sous la forme :

$$R_x = \frac{1}{T} \sum_{k=1}^T x_k x_k^H \quad (2.4)$$

où x_k est le vecteur signal échantillonné à l'instant k et T est le nombre d'échantillons.

2.2.2 Méthode MUSIC

La méthode *MUSIC* classique [7][9] est basée sur l'analyse de la matrice de covariance spatiale des signaux issus des capteurs. Elle suppose que la matrice de covariance des signaux des sources $R_s(t) = E[s(t) \ s^H(t)]$ est de rang plein P et que les vecteurs directionnels $a(\theta_1), a(\theta_2), \dots, a(\theta_P)$ sont linéairement indépendants. La matrice de transfert A est de rang plein P , donc

$$\text{rang}(A(\theta)R_sA^H(\theta)) = P \quad (2.5)$$

D'après la relation (2.5), $A(\theta) R_s A^H(\theta)$ a P valeurs propres strictement positives, et les autres ($N - P$) valeurs propres sont toutes égales à zéro.

La décomposition spectrale de la matrice de covariance R en éléments propres pour séparer le sous-espace signal du sous-espace bruit peut être exprimée sous la forme suivante :

$$R = AR_sA^H + \sigma^2 I \quad (2.6)$$

$$= U\Lambda U^H \quad (2.7)$$

$$= U_s\Lambda_sU_s^H + \sigma^2 U_bU_b^H \quad (2.8)$$

où Λ est la matrice diagonale des valeurs propres de la matrice R et U est la matrice formée des vecteurs propres correspondant aux valeurs propres de la matrice R classées par ordre décroissant.

On a donc :

$$\Lambda = \text{diag}[\lambda_1 > \lambda_2 > \dots > \lambda_p > \lambda_{p+1} = \lambda_{p+2} = \dots = \lambda_N = \sigma^2] \quad (2.9)$$

et $U = [U_s \ U_b]$ avec $U_s = [u_1, u_2, \dots, u_P], U_b = [u_{P+1} u_{P+2} \dots u_N]$, où U_s et U_b représentent respectivement les matrices des valeurs propres associées au sous espace signal et au sous- espace bruit .

Les vecteurs propres formant U_b correspondent aux plus petites valeurs propres et sont orthogonaux aux vecteurs colonnes de la matrice de transfert A :

$$\{u_{P+1} u_{P+2} \dots u_N\} \perp \{a(\theta_1), a(\theta_2), \dots, a(\theta_P)\}. \quad (2.10)$$

On a donc :

$$A^H U_b = 0. \quad (2.11)$$

Cette orthogonalité est due au fait que les vecteurs du sous-espace signal engendrent le même sous espace que les vecteurs colonnes de la matrice de transfert A et comme U_s est orthogonal à U_b , les colonnes de A sont donc orthogonales à celles de U_b . Pour estimer les directions d'arrivée des sources, le modèle du vecteur source $a(\theta)$ doit être connu. On a supposé que les sources sont placées en champ lointain, les ondes reçues sont alors des ondes planes et les capteurs utilisés sont identiques et de gain unité. L'orthogonalité des vecteurs directionnels avec les vecteurs propres du sous-espace bruit est alors caractérisée par une projection du sous-espace signal sur le sous-espace bruit (c'est pourquoi la méthode *MUSIC* s'appelle aussi algorithme de sous-espace projection et on cherche les valeurs de θ pour lesquelles le sous-espace bruit est orthogonal au sous-espace signal, ce qui correspond aux directions d'arrivée des ondes. La fonction spectrale angulaire obtenue par la méthode *MUSIC* permet de déterminer les valeurs de θ pour lesquelles cette fonction est maximale et elle est définie sous la forme suivante :

$$P_{\text{MUSIC}}(\theta) = \frac{1}{a(\theta)^H U_b U_b^H a(\theta)} \quad (2.12)$$

Il faut noter que $P_{\text{MUSIC}}(\theta)$ n'est pas un véritable spectre (c'est une mesure de la distance entre deux sous-espaces), elle nous donne des pics correspondant aux directions exactes d'arrivée des ondes mais ne nous renseigne pas sur la puissance des sources. Pour

que l'algorithme soit opérationnel il faut obligatoirement avoir un sous– espace bruit et un sous espace–signal. Puisque la dimension de l'espace total est égale à N , si on prend un sous–espace bruit de taille minimale 1, alors la dimension du sous–espace signal nous donne le nombre maximal de sources pouvant être détectées soit $N – 1$. La condition de fonctionnement de l'algorithme est donc :

$$P \leq N – 1 \quad (2.13)$$

Cette condition restreint le nombre minimum des capteurs supérieur au nombre total des sources, ainsi l'hypothèse initiale sur le nombre des sources P devient très importante et un test est utilisé afin de déterminer le nombre des sources présentes dans le milieu.

Des travaux de recherche visant à améliorer la méthode *MUSIC* en termes de temps de calcul, de performance et de robustesse ont été développé Dans [5], les auteurs proposent une décomposition en valeurs singulières généralisée de la matrice de covariance (*MUSIC – GSVD*). Cette méthode augmente la performance de localisation de sources sonores en présence d'un bruit. L'algorithme *MUSIC – GSVD* a été implémenté dans le logiciel *HARK*.

2.3 Modélisation de l'environnement de localisation

La localisation de sources sonores avec un robot mobile nécessite une modélisation de l'environnement. Ce modèle permettra d'interpréter les informations acoustiques de la scène auditive, d'intégrer les données acquises, puis d'estimer la position des sources.

L'environnement peut être modélisé par deux méthodes, avec chacune des outils de mise à jour associés :

- La grille d'occupation utilise une discrétisation de l'environnement suivant les deux axes d'un repère cartésien, avec un pas de discrétisation commun. L'état d'occupation des cellules est décrit par des probabilités. Les outils de mise à jour sont alors des équations permettant de fusionner des probabilités provenant de sources différentes et intégrant les informations acquises,
- La modélisation par primitives géométriques décrit l'environnement avec des objets géométriques, tels que les segments, et modélise la position de ces objets par des variables aléatoires gaussiennes. Les outils de mise à jour sont la distance de Mahalanobis et le filtre de Kalman.

Selon Thrun [10], les cartes construites à partir du principe de la grille d'occupation sont le paradigme dominant pour la modélisation de l'environnement en robotique mobile.

Une fois acquis, elles permettent différentes fonctions clés nécessaires à la navigation de robots mobiles, telles que la localisation, la planification de chemin, l'évitement de collisions, ce paradigme de grille d'occupation a été appliqué avec succès de nombreuses manières différentes. Par exemple, certains systèmes utilisent des cartes locales, pour planifier des chemins sans collision ou pour identifier des caractéristiques de l'environnement pour la localisation.

Les cartes d'occupation ont été construites en utilisant des capteurs sonar, des télémètres laser, et la stéréovision. Alors que la plupart des algorithmes de cartographie d'occupation existants utilisent des cartes en deux dimensions[10][11][12][13], certains ont effectivement développé en trois dimensions des cartes volumétriques[14]. Plusieurs auteurs ont étudié la construction de ces cartes pour la localisation simultanée du robot et l'exploration de l'environnement.

Dans [15], l'auteur propose de localiser les sources sonores présentes dans l'environnement en deux étapes :

- La première étape consiste à l'exploration de l'environnement par un robot mobile, qui est chargé d'enregistrer les événements auditifs intéressants à l'aide d'un réseau de microphones. les données enregistrées permettent de localiser les sources sonores les plus pertinentes et par la même occasion de mettre à jour la grille d'occupation qui modélise la scène auditive,
- La seconde étape consiste alors à étudier les sources sonores détectées et leurs environs afin de construire des modèles de champ proche de chaque source sonore.

Cependant la représentation de la scène auditif par une grille d'occupation classique suppose que la source sonore s'étale sur plusieurs cellules, cela est dû à l'indépendance des cellules entre elles. Dans l'étape deux y a aucune preuve d'optimalité quant au déplacement du robot dans la grille.

2.4 Conclusion

Pour cartographie un environnement sonore avec un robot mobile, un ensemble d'outils sont à définir et à étudier :

- Un modèle de l'environnement capable d'interprète, d'intègre les données des capteurs et de prendre compte des caractéristiques des sources sonores.
- Une méthode de localisation de source sonores robuste au bruit du robot mobile, et d'estimer la direction d'arrivée *DOA* (*Direction Of Arrival*) des sources sonores en temps réel.

- Une stratégie de déplacement de robot optimale, tenant compte des positions des sources sonores.

Chapitre 3

Approche Proposée

On considère un robot mobile intelligent qui évolue dans une pièce fermée. Ce milieu est dense et relativement structuré. On souhaite définir un modèle géométrique de l'environnement du robot, afin de localiser les sources sonores présentes dans cet environnement. Il existe différents modèles pour représenter l'environnement d'un robot mobile (le modèle basé sur les primitives, basé sur des grilles, hybride...). On se focalisera dans ce document sur un modèle basé sur les grilles d'occupation.

3.1 Grille d'occupation

Pour assurer la tâche de localisation des sources sonores par le robot mobile, le robot doit construire une représentation de l'environnement. Parmi les modèles de représentation de l'environnement, des approches discrètes sont couramment utilisées en robotique, notamment au travers de la représentation par "grilles d'occupation" [Elfes, 1989c]. Ce formalisme permet de représenter l'environnement sans faire d'hypothèse sur la géométrie des éléments présents.

La grille d'occupation 2D utilise un maillage structuré pour partitionner la plan horizontal de l'environnement en cellules. Ainsi la grille ressemble à une vue de dessus de la scène.

Différentes formulations existent, utilisant soit un modèle de capteur direct soit un modèle inverse et considérant un environnement soit statique soit dynamique. Dans ce type d'approche, on utilise le principe d'inférence bayésienne pour estimer la probabilité d'occupation des cellules par une source sonore.

Ceci est basé sur le théorème de Bayes :

$$P[A | B] = \frac{P[B | A]P[A]}{P[B]}.$$

L'utilisation de ce formalisme avec une structure de grille a été proposée par Elfes dans [Elfes, 1989c, Elfes, 1989a]. Chaque cellule i de la grille d'occupation G_O contient une probabilité d'occupation qui est liée à la possibilité de contenir un objet visuel ou une source sonore. Dans l'approche originelle, l'état d'occupation de chaque cellule est évalué indépendamment. Dans notre cas les états de toutes les cellules de la grille sont dépendants les uns des autres. Cela est dû à plusieurs raisons :

- On veut estimer la position d'une source sonore unique (alors que visuellement on a toujours plusieurs points).
- Si le robot mobile ne se déplace pas, le fait d'estimer chaque cellule indépendamment des autres nous estimera qu'une source sonore s'étale sur une droite alors qu'une source sonore est un objet ponctuel.
- Cela nous permet de calculer une fonction d'entropie qui nous permettra d'estimer la quantité d'information contenue par une grille et de l'utiliser comme objectif pour la commande.

On pose G_O comme étant une grille d'occupation. On notera i une cellule de la grille d'occupation G_O avec les indices $\{x_i, y_i, z_i\}$. L'occupation d'une cellule de la grille par une source sonore est estimée à partir des mesures (Kinect, Laser...). On notera $m_{1:T}$ les mesures de l'instant 1 jusqu'à T . Chaque mesure implique une information sur l'occupation des cellules de la grille par des sources sonores. La grille d'occupation G_O formalise la probabilité d'occupation d'une cellule i par une source sonore sachant la mesure $m_{1:T}$, les positions et l'orientation $p_{1:T}$ du robot mobile. En notant x la position de la source sonore,

$$G_0(i) = P[x = i \mid m_{1:T}, p_{1:T}] \quad (3.1)$$

Les grilles sont définies avec une très grande dimension, ce qui nous permet d'avoir une très grande précision.

Pour le calcul de la probabilité *a posteriori* (3.1), on se propose d'utiliser le formalisme probabiliste, qui a été le premier formalisme de gestion des incertitudes utilisé dans les grilles d'occupation.

3.2 Modèle de capteur

Les différents capteurs extéroceptifs(Kinect,Laser...) utilisés fournissent des informations sur l'espace autour du robot. Ces capteurs sont des dispositifs qui permettent de traduire des propriétés physique de l'environnement en données exploitables. On considère ici un capteur Kinect constitué de quatre microphones positionnés sur un axe horizontal.

Comme on s'intéresse à la position des sources sonores, il faut être capable de traduire les grandeurs fournies par les capteurs en informations sur la confiance d'occupation. Pour cela on définit une fonction que l'on nommera le modèle de capteur. Le modèle peut être déterminé empiriquement (statistiquement) ou fixé manuellement.

Pour simplifier, nous faisons les hypothèses suivantes :

- On suppose que toutes les sources sonores qu'on cherche à localiser sont à la même hauteur que les microphones de la Kinect.
- La grille d'occupation est représentée sur la même plan horizontal que l'axe des microphones de la Kinect.
- On notera i cellule de la grille dont les indices sont x_i, y_i au lieu de x_i, y_i, z_i .
- On suppose que la Kinect et le robot mobile sont solidaire et que la position de la Kinect $p_t(x_t, y_t, \theta_t)$ est caractérisée par sa position linéaire (x_t, y_t) et par son orientation (θ_t) par rapport à l'axe y de la pièce.

On suppose que les mesures sont des angles horizontaux (*azimuths*) estimés par l'algorithme *MUSIC – GSVD* [Okuno,2013]. Ces mesures ne prennent pas en compte la distance de la source dans l'espace mais uniquement sa direction d'arrivée.

Notre modèle de capteur est défini par :

$$p[m_t | x = i, p_t] = p[m_t | d_{i_t}, \alpha_{i_t}] \quad (3.2)$$

où :

- m_t est la mesure fournie par l'algorithme *MUSIC – GSVD* à l'instant t dans le repère du robot.
- d_{i_t} est la distance à laquelle se trouve une cellule i à l'instant t par rapport au capteur.
- α_{i_t} est la direction d'arrivée de la cellule i à l'instant t par rapport au capteur.

Pour estimer le modèle de capteur, une expérience a été effectuée par E.Vincent sous Matlab.

3.2.1 Description de l'expérience

Pour cette expérience, on suppose que la scène sonore comprend une seule source sonore de parole avec une puissance à 30 cm de 70 dB SPL superposée à un bruit de fond diffus de puissance 45 dB SPL. Le spectre du bruit, le rapport signal-à-bruit et le temps de réverbération approximativement correspondent à ceux observés dans l'appartement intelligent d'Inria Nancy. Les distances entre microphones correspondent à celles du capteur Kinect. Sans prendre en compte les *HRTFs* de la Kinect¹ (qui ne jouent un rôle qu'en haute fréquence) on suppose que les microphones sont omnidirectionnels. La localisation est effectuée par l'algorithme *MUSIC – GSVD* sur des trames de 64 ms toutes les 16 ms.

$p[m_t | d_{it}, \alpha_{it}]$ est la probabilité que la direction estimée par *MUSIC – GSVD* soit égale à m_t lorsque la vraie direction et la distance de la source sont égales à α_{it} et d_{it} . Les variables m_t et α_{it} varient entre 0° et 359° , où 0° correspond à une source alignée avec l'axe des microphones (sachant que la Kinect dispose de quatre microphones). Les angles sont mesurés dans le sens trigonométrique par rapport à cet axe. La variable d_{it} prend 5 valeurs : 50 cm, 1 m, 1.50 m, 2 m ou 3 m.

Pour une valeur quelconque de m_t (en $^\circ$), α_{it} (en $^\circ$) et d_{it} (en m), on peut donc calculer la probabilité $p(m_t | d_{it}, \alpha_{it})$ en interpolant linéairement entre $p[m_t | \alpha_{it}, d_1]$ et $p[m_t | \alpha_{it}, d_2]$ pour les deux distances d_1 et d_2 les plus proches.

$$p[m_t | \alpha_{it}, d] = \frac{p[m_t | \alpha_{it}, d_2] - p[m_t | \alpha_{it}, d_1]}{d_2 - d_1} (d - d_1) + p[m_t | \alpha_{it}, d_1] \quad \forall d \in [d_1, d_2]$$

1. *HRTFs* : fonction de transfert relative à la Kinect (*head – related transfer function*) qui caractérise les transformations apportées aux ondes sonores par le corps de la Kinect

3.2.2 Résultat de l'expérience

Les résultats de l'expérience sont illustré dans les figures 3.1 à 3.5.

La distribution de probabilité est concentrée autour du bon angle et de son symétrique

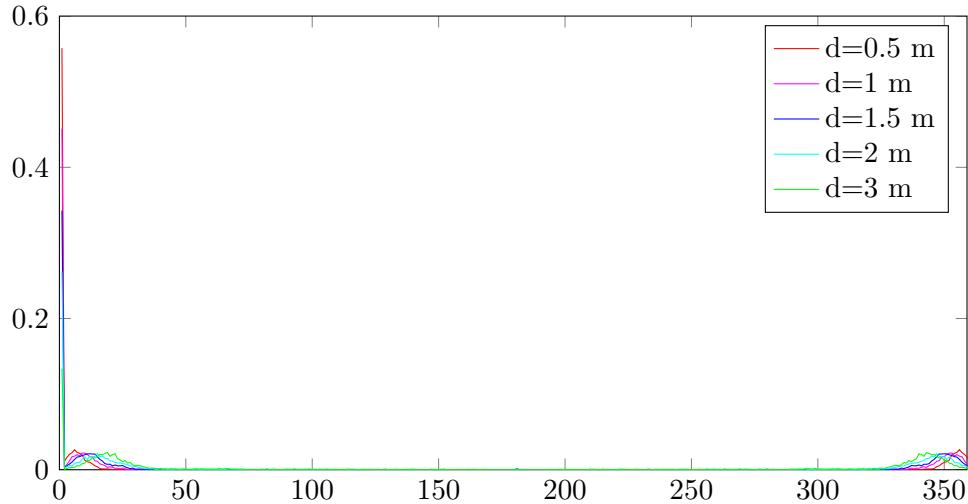


FIGURE 3.1: Modèle de capteur $p[m_t | d, \alpha]$ pour une source de direction d'arrivée $\alpha = 0^\circ$, pour différentes distances d.

par rapport à l'axe des capteurs à faible distance, mais elle s'étale et des faux pics apparaissent à 0° et 180° à grande distance en raison de l'accroissement relatif du bruit.

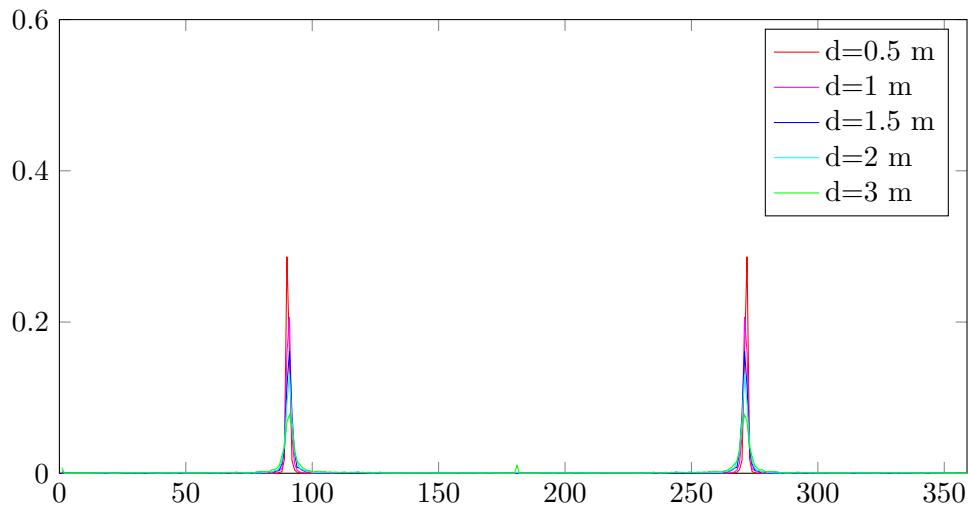


FIGURE 3.2: Modèle de capteur $p[m_t | d, \alpha]$ pour une source de direction d'arrivée $\alpha = 90^\circ$, pour différentes distances d.

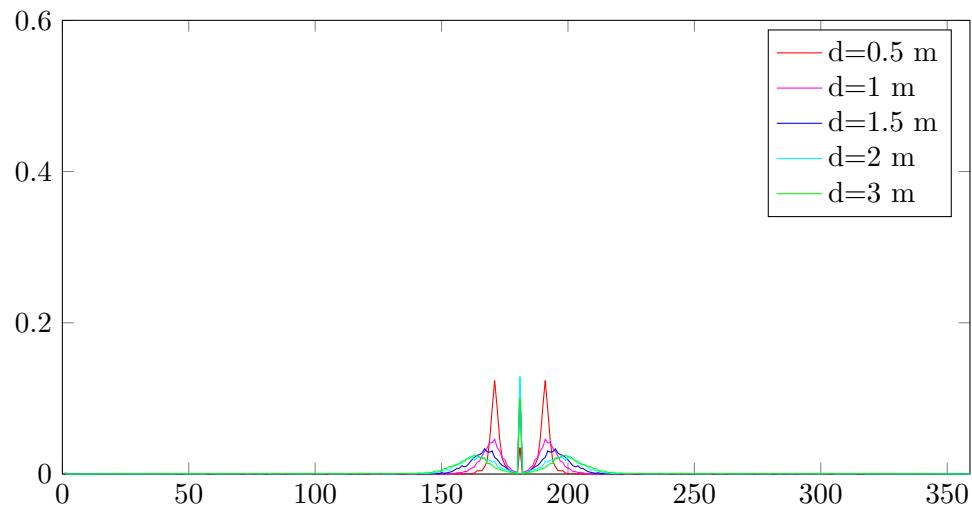


FIGURE 3.3: Modèle de capteur $p[m_t | d, \alpha]$ pour une source de direction d'arrivée $\alpha = 170^\circ$, pour différentes distances d.

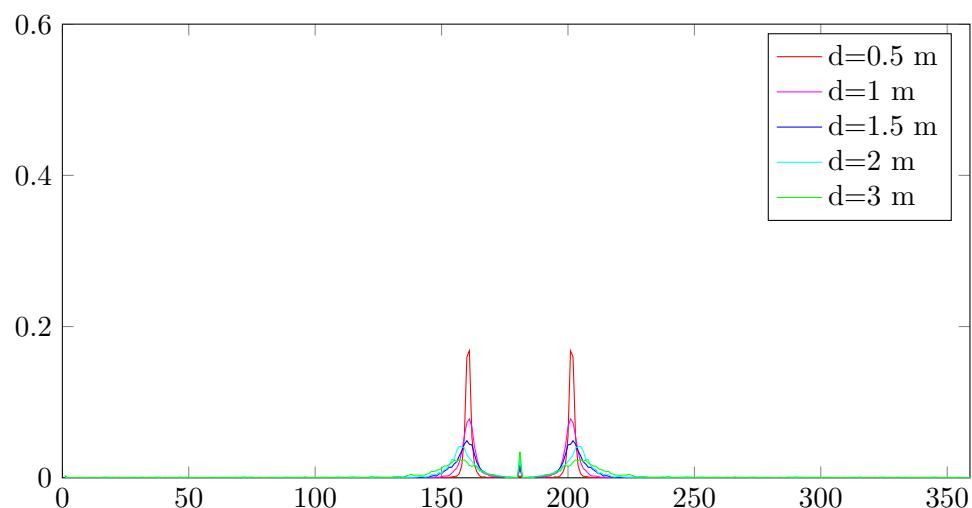


FIGURE 3.4: Modèle de capteur $p[m_t | d, \alpha]$ pour une source de direction d'arrivée $\alpha = 200^\circ$, pour différentes distances d.

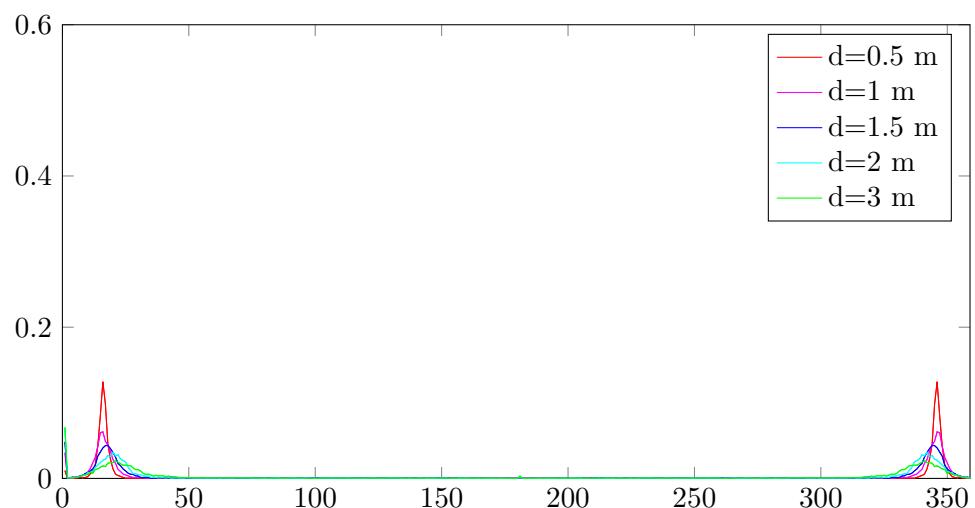


FIGURE 3.5: Modèle de capteur $p[m_t | d, \alpha]$ pour une source de direction d'arrivée $\alpha = 345^\circ$, pour différentes distances d .

3.3 Mise à jour de la grille d'occupation

Pour chaque cellule i , on cherche la probabilité d'occupation *a posteriori*

$$P[x = i | m_{1:T}, p_{1:T}]$$

où

- $m_{1:T} = (m_1, \dots, m_T)$ représente l'ensemble des mesures obtenues jusqu'à l'instant T,
- $p_{1:T} = (p_1, \dots, p_T)$ représente l'ensemble des positions et orientations du robot jusqu'à l'instant T ,
- i est une cellule de la grille ,
- x est la position de la source sonore.

La règle de Bayes fournit l'équation de mise à jour suivante :

$$P[x = i | m_{1:T}, p_{1:T}] = \frac{P[m_T | x = i, p_T]P[x = i | m_{1:T-1}, p_{1:T-1}]}{\sum_{i'} P[m_T | x = i', p_T]P[x = i' | m_{1:T-1}, p_{1:T-1}]} \quad (3.3)$$

Cette équation permet de mettre à jour la grille de façon itérative à chaque fois qu'une nouvelle mesure m_t est disponible. Cette mise à jour consiste à calculer la probabilité à posteriori $P[x = i | m_{1:T}, p_{1:T}]$, à partir de la probabilité à priori $P[x = i | m_{1:T-1}, p_{1:T-1}]$ donnée par l'itération précédente et de la vraisemblance $P[m_T | x = i, p_T]$ de la mesure.

3.4 Commande du robot

3.4.1 Définition de la récompense

L'équation (3.3) permet de localiser les sources dans une certaine mesure même avec un déplacement aléatoire du robot. Le but de la commande est de localiser les sources le plus vite possible (ou autrement dit du mieux possible dans un laps de temps fini). Pour localiser les sources sonores, on se propose de déplacer le robot d'une manière déterministe (contrôlée) autour des sources sonores présentes sur la grille d'occupation. Cela nous amène à un problème de planification de trajectoire et de localisation du robot.

Dans notre cas pour simplifier on suppose que la position du robot est parfaitement connue, et cela grâce à des capteurs proprioceptifs (odométrie) et des capteurs extéroceptifs (vision,laser). Il nous reste le problème de planification de la trajectoire.

Pour un ensemble de mesures données on quantifie l'information portée par la grille par son entropie :

$$H[x \mid m_{1:T+K}, p_{1:T+K}] = - \sum_i P[x = i \mid m_{1:T+K}, p_{1:T+K}] \log P[x = i \mid m_{1:T+K}, p_{1:T+K}] \quad (3.4)$$

L'équation (3.4) représente l'information finale (avec un signe (-) et à une constante près).

Pour planifier, il faut être capable de prédire la valeur de l'entropie consécutive à un déplacement future du robot mobile. Pour cela on utilise l'espérance sur les mesures futures. Une première approche consiste à calculer le gain total d'information sur une longue période. Cela nous permettra dans le cas idéal de planifier la position finale et les positions intermédiaires du robot, qui nous informent le plus sur la position des sources sonores. On appellera cette première approche le critère idéal, défini par :

$$\mathbb{E}_{m_{T+1:T+K}}[H[x \mid m_{1:T+K}, p_{1:T+K}]] = \sum_{m_{T+1:T+K}} P[m_{T+1:T+K} \mid m_{1:T}, p_{1:T+K}] H[x \mid m_{1:T+K}, p_{1:T+K}] \quad (3.5)$$

où :

$$P[m_{T+1:T+K} \mid m_{1:T}, p_{1:T+K}] = \sum_i P[x = i \mid m_{1:T}, p_{1:T}] \prod_{k=1}^K P[m_{T+k} \mid x = i, p_{T+k}] \quad (3.6)$$

Ce critère Idéal n'est pas calculable en pratique.

Une deuxième approche se base sur le calcul du gain d'information par trame, ce qui veux dire que pour chaque position susceptible de se trouver sur la trajectoire du robot à un instant donné $T + k$ on estime le gain d'information apporté par cette position par rapport à la position initiale, sans prendre en compte les positions intermédiaires. Cela nous amène à un problème classique de programmation dynamique. On appellera cette deuxième approche le critère approché, défini par :

$$c_{p_{T+k}} = \mathbb{E}_{m_{T+k}}[H[x \mid m_{1:T}, m_{T+k}, p_{1:T}, p_{T+k}]] \quad (3.7)$$

$$= \sum_{m_{T+k}} P[m_{T+k} \mid m_{1:T}, p_{1:T}, p_{T+k}] H[x \mid m_{1:T}, m_{T+k}, p_{1:T}, p_{T+k}] \quad (3.8)$$

où :

$$\begin{aligned} H[x \mid m_{1:T}, m_{T+k}, p_{1:T}, p_{T+k}] &= - \sum_i P[x = i \mid m_{1:T}, m_{T+k}, p_{1:T}, p_{T+k}] \\ &\quad \log P[x = i \mid m_{1:T}, m_{T+k}, p_{1:T}, p_{T+k}] \end{aligned} \quad (3.9)$$

$$P[m_{T+k} \mid m_{1:T}, p_{1:T}, p_{T+k}] = \sum_i P[x = i \mid m_{1:T}, p_{1:T}] P[m_{T+k} \mid x = i, p_{T+k}] \quad (3.10)$$

3.4.2 De la programmation dynamique pour la planification de trajetoire

La figure (3.6) illustre les différentes positions possibles (pour une orientation donnée) du robot mobile à chaque trame.

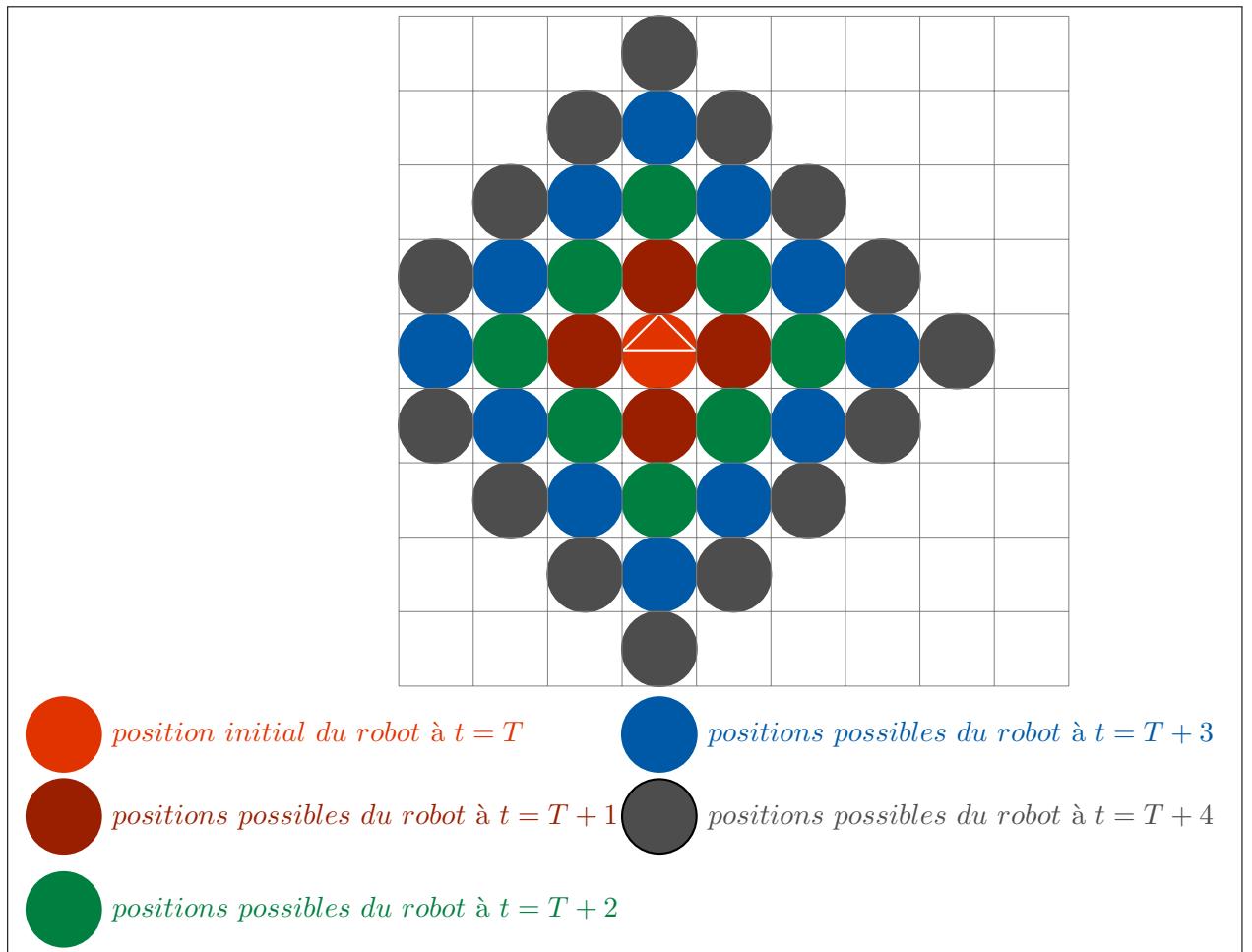


FIGURE 3.6: Positions possibles du robot mobile p_{T+k} (pour une orientation donnée)
à chaque instant $T + k$ dans le plan 2D

Pour simplifier le problème de planification de trajectoire on décrit le déplacement du robot à l'aide d'un graphe séquentiel qui nous simplifiera les calculs de la trajectoire du robot. la figure(3.7) représente le graphe de séquence associé à la figure (3.6).

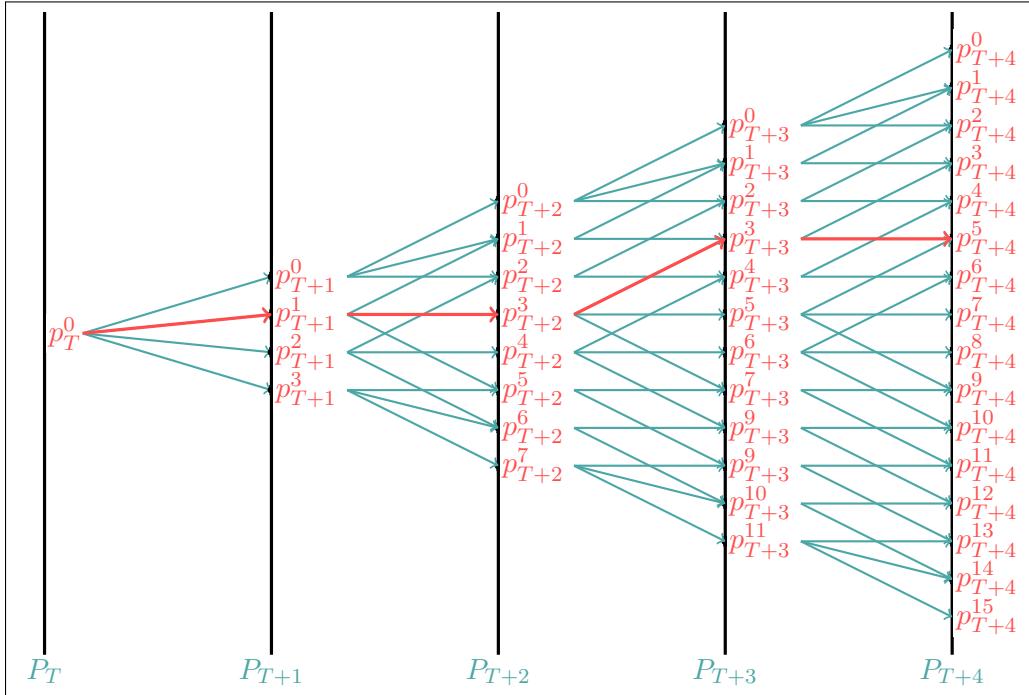


FIGURE 3.7: Graphe séquentiel associé au déplacement du robot mobile

Reprendons la figure (3.6) qui représente les positions possibles du robot mobile dans la grille d'occupation. Le graphe associé peut devenir séquentiel si l'on divise les chemins en trames (les différentes positions possibles au robot sur chaque trame sont disposées en colonnes successives), et si l'on impose un sens de déplacement du robot alors chaque chemin est une séquence d'arcs allant d'un sommet d'une colonne à un sommet de la colonne suivante. Étudions ainsi les chemins allant des sommets d'entrée (de départ) aux sommets de sortie (finaux). Dans notre cas les sommets de départ représentent les positions possibles à l'instant T et les états finaux représentent les positions possible à l'instant $T + K$ (K horizon de planification préalablement fixé).

Le cheminement le long d'un arc (p_{T+h}^l, p_{T+k}^r) correspond à la décision si l'on se trouve en p_{T+h}^l d'aller de p_{T+h}^l à p_{T+k}^r . Le chemin $p_T \rightarrow p_{T+4}^n$ est le résultat de 4 décisions successives, par exemple la séquence $p_T^0, p_{T+1}^1, p_{T+2}^2, p_{T+3}^3, p_{T+4}^4$ correspond aux choix : emprunter l'arc (p_T^0, p_{T+1}^1) , puis l'arc (p_{T+1}^1, p_{T+2}^2) , l'arc (p_{T+2}^2, p_{T+3}^3) et enfin l'arc (p_{T+3}^3, p_{T+4}^4) . Une fois qu'on a défini notre problème de planification sous forme d'un graphe séquentiel, il est facile d'appliquer la méthode de programmation dynamique et ainsi de trouver la stratégie optimale (chemin optimal).

3.4.3 Recherche du chemin de valeur optimale

La recherche du chemin optimal s'appuie sur l'algorithme de *Bellman*. Les programmes les plus simples sont ceux où l'avenir est déterminé. L'intérêt de l'algorithme est évident quand le caractère combinatoire du problème amène à comparer un très grand nombre de politiques.

Étant donné un graphe $G = (X, U)$, on associe à tout arc $u = (x_0, x_1)$ $c(u)$ appelé coût de cet arc. Dans notre cas le coût c est associé à chaque position et orientation possible p_{T+k} du robot mobile.

Le coût d'un chemin sera alors la somme des coûts des arcs définissant ce chemin. On peut alors se poser la question : quel est le chemin de coût minimal conduisant le robot mobile du sommet p_T à un sommet p_{T+K} .

$$\min_{p_{T:T+K}} \sum_{k=1}^K c_{p_{T+k}} \quad (3.11)$$

Une première méthode de recherche consisterait à étudier tous les chemins existants entre le sommet d'entrée p_{T+1} et un sommet de sortie p_{T+K} et en établira le classement en fonction de leurs coûts.

Mais l'application de cette méthode devient beaucoup trop longue quand le nombre de chemins $p_{T+1} \rightarrow p_{T+K}$ est élevé. Il vaut mieux utiliser alors une deuxième approche basée sur le théorème d'optimalité de Bellman.

Ce théorème s'énonce ainsi : un chemin optimal ne peut être formé que de chemins partiels optimaux. Il peut se démontrer facilement par l'absurde.

Ce résultat fournit un algorithme de recherche du chemin de valeur optimale dans un graphe quelconque. La mise en œuvre de cet algorithme est particulièrement simple dans le cas des graphes séquentiels.

Considérons le graphe séquentiel (3.7) dont les sommets (les positions possibles du robot) sont répartis en sous-ensembles $P_T, P_{T+1} \dots P_{T+K}$. Un chemin optimal de P_{T+k} à P_{T+K} ne peut être formé que de chemins partiels optimaux de P_{T+k+1} à P_{T+K} . Cette propriété permet l'utilisation d'un algorithme "remontant" (vers les valeurs décroissantes de k) pour trouver le(ou les) chemin(s) à valeur optimale entre entrée et sorties du graphe.

Soit la fonction de K variable à optimiser :

$$F(p_{T+1}, p_{T+3} \dots p_{T+K}) = \sum_{k=1}^K c_{p_{T+k}} \quad (3.12)$$

où $c_{p_{T+k}}$ est le coût de la position p_{T+k} à l'instant $T + k$.

On cherche $f_{T+2}(p_{T+2})$, la valeur optimale de $c_{p_{T+1}} + c_{p_{T+2}}$ c'est à-dire :

$$f_{T+2}(p_{T+2}) = \min_{p_{T+1}} (c_{p_{T+1}} + c_{p_{T+2}}) \quad (3.13)$$

Ensuite, on cherche la valeur optimale $f_{T+3}(p_{T+3})$

$$f_{T+3}(p_{T+3}) = \min_{p_{T+2}} (f_{T+2}(p_{T+2}) + c_{p_{T+3}}) \quad (3.14)$$

En continué ensuite jusqu'à trouver $f_{T+k}(p_{T+K})$. Ainsi, l'optimum recherché est obtenu avec la formule

$$F(p_{T+1}, \dots, p_{T+K}) = \min_{p_{T+K}} f_{T+K}(p_{T+K}) \quad (3.15)$$

Chapitre 4

Évaluation expérimentale

Afin d'évaluer la procédure de localisation de sources sonores proposée dans le chapitre³ dans des conditions réelles, des expériences ont été réalisées dans l'appartement Intelligent d'Inria Nancy. Les résultats obtenus sont comparés au résultats obtenus par la procédure aléatoire. Une description détaillée du protocole expérimental et des deux procédures ainsi des mesures d'évaluation sont fournies. Ensuite les résultats seront présentés.

4.1 Protocole expérimental

L'appartement intelligent du centre de recherche Inria Nancy [4.1](#) est une pièce dans laquelle on a réalisé l'ensemble de nos expériences. Une source de parole statique émise par un haut parleur et le robot mobile (Turtlebot 2) sont présents dans cette pièce.

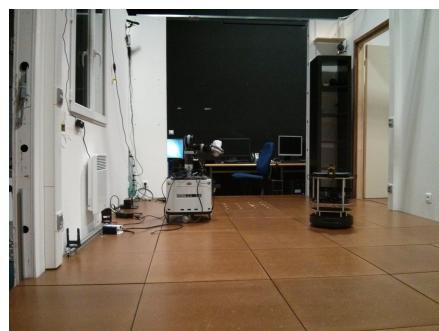


FIGURE 4.1: Photo de l'appartement intelligent du centre de recherche INRIA de Nancy

L'acquisition du signal émis par la source de parole est effectuée par les quatre microphones du capteur Kinect (Fig.[4.3a](#)) fixé sur le TurtleBot 2(Fig.[4.2](#)). Le signal enregistre

par la Kinect est traité par le PC embarqué sur le robot mobile. Le capteur Kinect et le Haut parleur sont placés à la même hauteur.



FIGURE 4.2: Photo du robot expérimental TurtleBot 2

Pendant l'expérience, les positions du robot mobile par rapport au repère fixe lié à l'appartement sont acquises avec un système de capteurs proprioceptifs (odométrie, gyroscope) de capteurs extéroceptifs (Laser (Fig. 4.6b), cameras de la Kinect) de haute définition donnant une erreur de position inférieure à dix centimètres. Les positions du haut-parleur sont acquises par rapport au repère fixe lié à l'appartement par un laser d'une grande précision (± 1 mm).



FIGURE 4.3: Capteurs du TurtleBot

La partie de l'appartement où sont réalisés nos expériences est modélisée par une grille d'occupation (Fig. 4.4) de haute définition ($5\text{cm} \times 5\text{cm}$). le haut-parleur et le robot mobile sont représenté par une cellule de la grille.

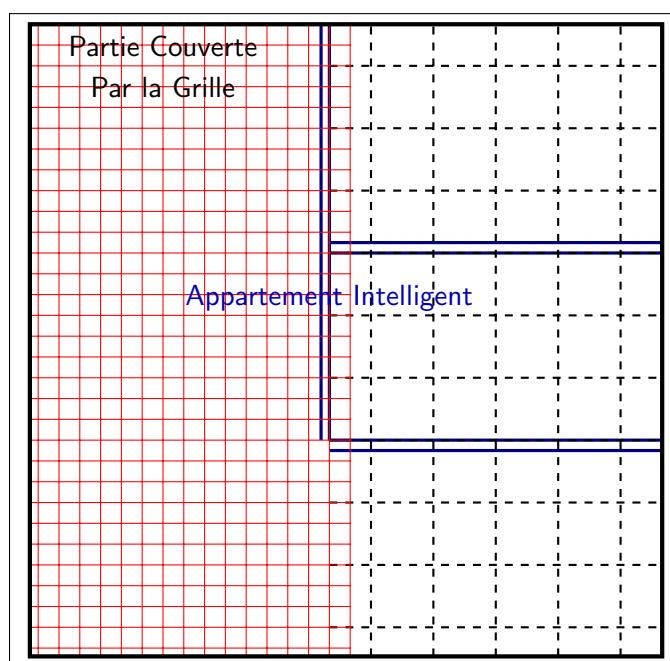


FIGURE 4.4: Couverture de la grille par rapport à la pièce

Angle ($^{\circ}$)	45		135		90	
distance (m)	2.4	1.2	2.4	1.2	2.4	1.2

TABLE 4.1: Récapitulatifs des configurations expérimentées

Pour nos expériences on considère deux variables :

- la distance euclidien de la source par rapport à l'origine du repère fixe où se trouve initialement le robot,
- la position angulaire de la source sonore par rapport au repère fixe.

Les détails de ces configurations sont données dans le tableau(4.1) puis illustré par la figure(4.5).

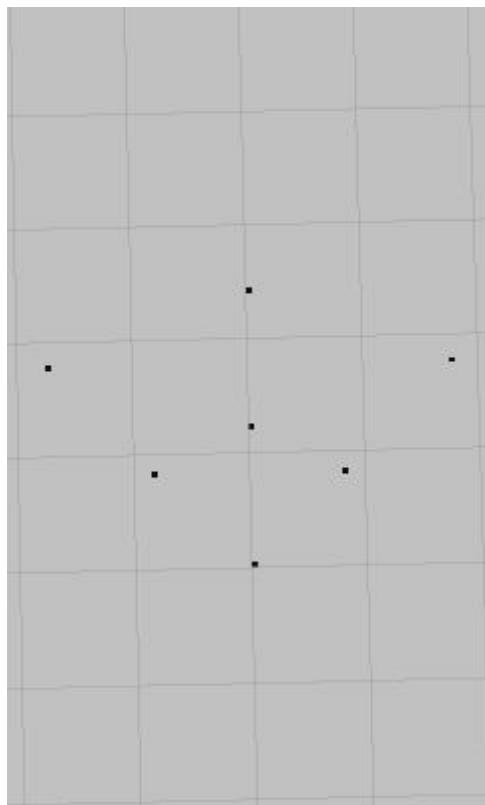


FIGURE 4.5: Différentes configurations de la source sonore dans la grille d'occupation

4.2 Procédure expérimentale

Pour nos expériences, on a mis en place deux procédures de localisation de la source sonore :

- une procédure automatisée basée sur l'approche proposée dans le chapitre (3), Le robot planifié son chemin d'une manière automatique, puis se déplace d'une manière complètement autonome,

- une procédure aléatoire où les déplacements du robot est complètement aléatoire.
- Pour chaque expérience, le robot mobile parcourt une distance près déterminer, pour une source sonore se trouvant à 1,2 m de l'origine du repère fixe, le robot parcourt une distance de 2,1, pour une source se trouvant à une distance de 2,4 m le robot se déplace de 3,0 m. avec un pas commun à 30 cm . c'est à dire que le robot se déplace de 30 cm à chaque itération jusqu'à atteindre la distance total (3 m pour une source sonore à 2,4 m et 2,1 à 12 positions pour une source à 1,2m ce qui revient à 8 positions). le robot mobile effectue une seul et unique mesure à chaque position.

4.3 Mesures d'évaluation

Les deux procédures proposés ci-dessus ont été évalués selon deux mesures :

- Calcul de l'erreur en position de la source sonore dans la grille. En notant par c_{estime} de coordonnées (x_{estime}, y_{estime}) cellule à plus forte probabilité de la grille et par c_{reel} les coordonnées exacte de la source (x_{reel}, y_{reel}) . la mesure d'erreur sur la position est calculée par la formule suivante :

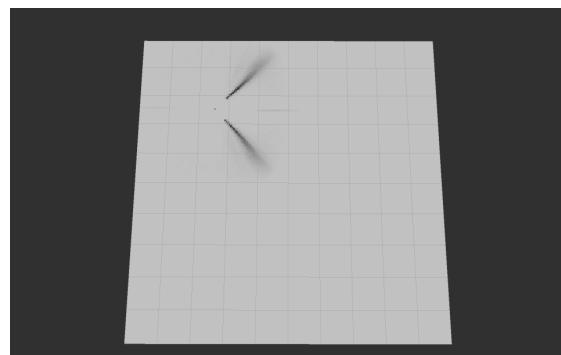
$$erreur(m) = \sqrt{(x_{estime} - x_{reel})^2 + (y_{estime} - y_{reel})^2} \quad (4.1)$$

- Le nombre de positions qu'il a fallu pour localiser la source sonore.

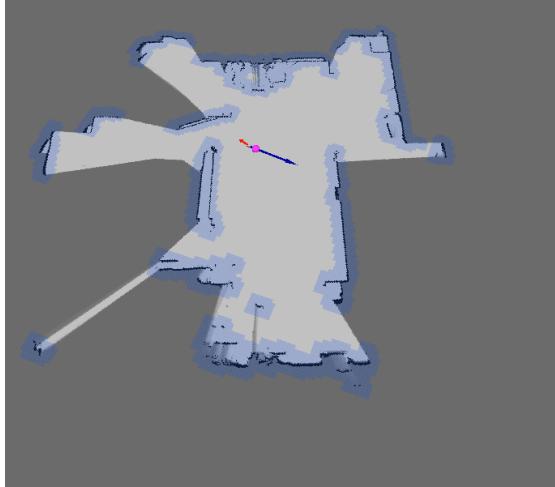
4.4 Résultat

On a développé les deux procédures avec le langage de programmation C++ en 150 ligne, les différentes tâches du robot sont commandées a partir du middleware ROS, qui dispose d'une interface graphique *rviz* permettant de visualiser les données des capteurs et les résultats de la localisation.

À partir des mesures fournies par l'algorithme *MUSIC-GSVD*, nous pouvons modéliser l'environnement sonore du robot mobile. Les deux procédures diffèrent seulement sur la façon de générer la trajectoire du robot. L'odomètre du robot demeure peu fiable, pour plus de précision on a recours au données laser et vision pour estimer la position du robot mobile.



(A) Carte de l'environnement sonore



(B) Carte basée sur les données laser et visuel du robot

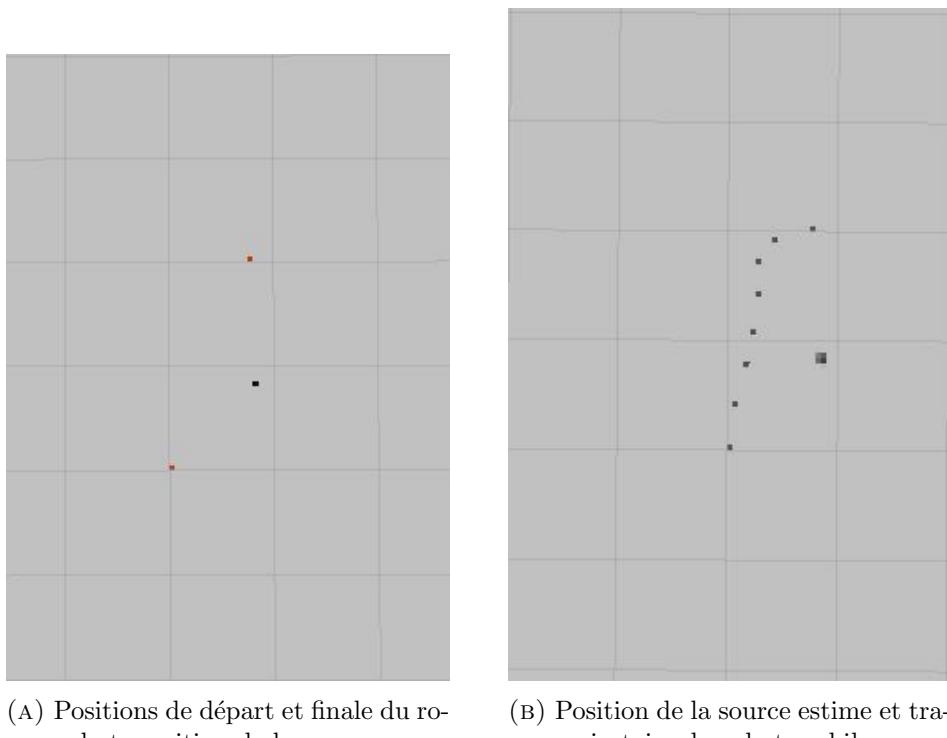
FIGURE 4.6: Modélisation de l'environnement du robot mobile

Nous avons donc : deux cartes une carte pour modéliser l'environnement sonore(Fig.4.6a), une deuxième pour modéliser et gérer le déplacement du robot mobile(Fig.4.6b). Cette

dernière est mise à jour à partir du laser et de la camera Kinect. La commande du robot mobile est générée avec un module ROS(Robot Operating System). Ce module génère la trajectoire optimale a partir d'une position fournie par une des deux procédures de localisation de la source sonore. Dans le cas de la procédure aléatoire, les positions du robot sont commandées d'une façon aléatoire. Par contre dans la procédure automatisée les positions du robot sont générées à l'aide de l'algorithme de programmation dynamique développé dans le chapitre 3.

Les résultats de la procédure automatique sont illustrés dans les figures 4.7 à 4.12.

Le temps de calcul de la trajectoire optimale est plus important que celui de la



(A) Positions de départ et finale du robot, position de la source (B) Position de la source estimée et trajectoire du robot mobile

FIGURE 4.7: Résultats de la trajectoire optimale pour une source sonore à 1.20 m et 45°

procédure aléatoire, car a chaque pas effectué par le robot mobile, il se doit de calculer la position suivante. Ce qui n'empêche pas cette procédure d'être plus rapide dans la localisation de la source sonore, car au bout de quelques postions visitées par le robot l'erreur sur la position de la source sonore diminue d'une manière considérable, jusqu'à ce que l'erreur soit inférieure à 0,10 m. Ce qui n'est pas le cas de la procédure aléatoire où l'erreur sur la position de la source est en moyenne supérieure à 0,25 m.

Durant les expériences on a remarqué que les deux procédures dépendent de l'orientation du robot à l'état initial. À la différence de la procédure aléatoire qui n'a aucun

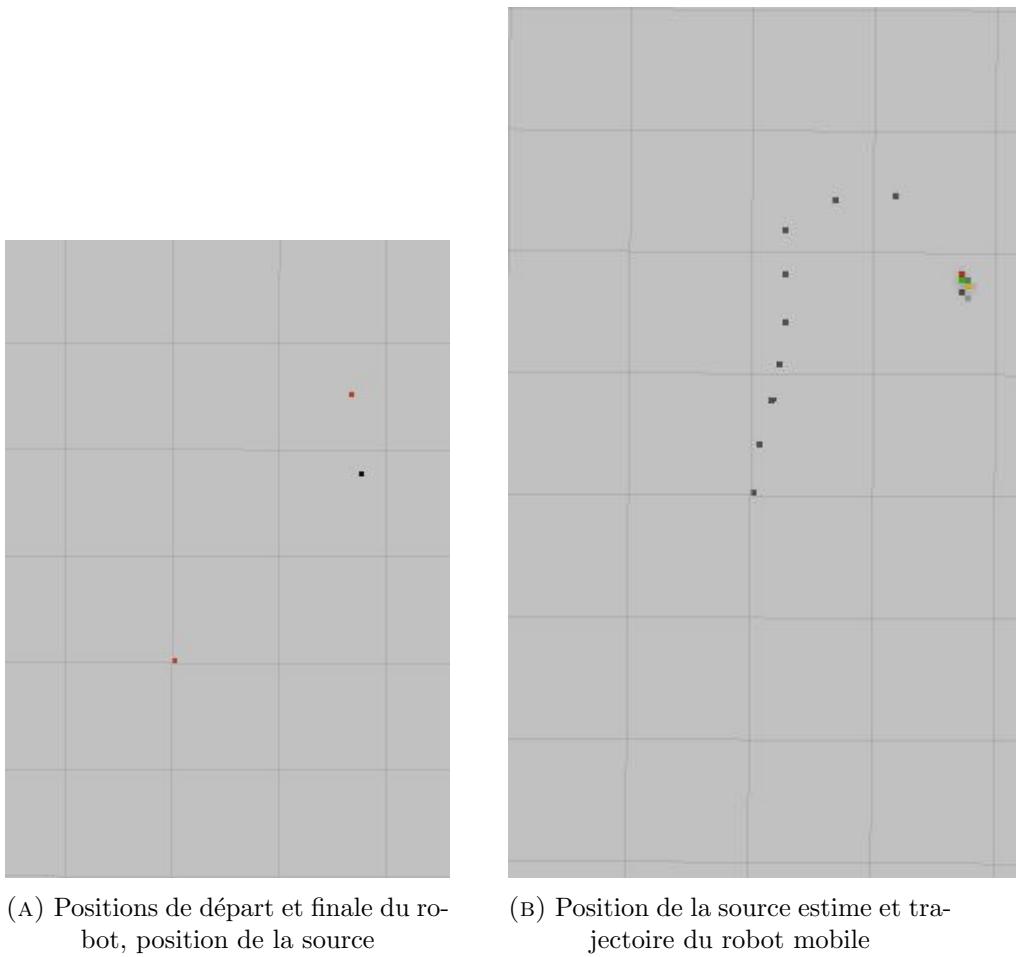
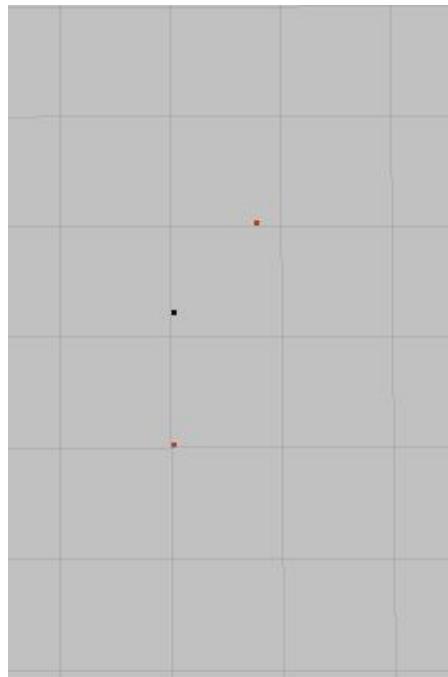
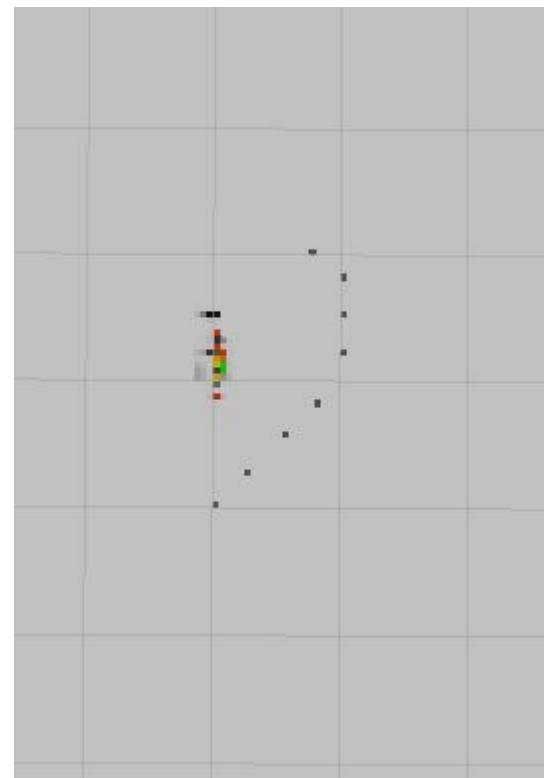


FIGURE 4.8: Résultats de la trajectoire optimale pour une source sonore à 2.40 m et 45°

contrôle sur l'orientation du robot, la procédure automatique définit l'orientation optimale du robot ce qui lui permet d'être plus précise et plus efficace en terme de temps de localisation malgré un temps de calcul relativement élevé. Lors des expériences on a aussi remarqué que dans certain cas les déplacement du robot ne sont pas égaux ($0, 30 \pm 0.10$), cela est dû à l'erreur d'estimation de la position du robot par le laser.

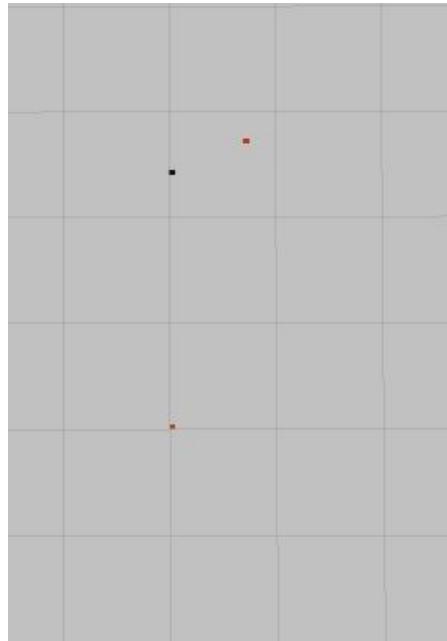


(A) Positions de départ et finale du robot, position de la source

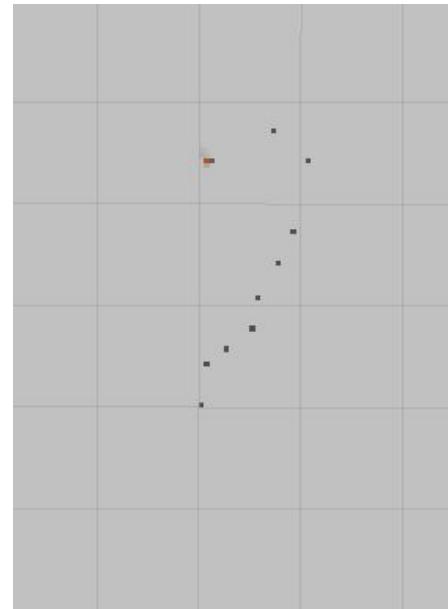


(B) Position de la source estimée et trajectoire du robot mobile

FIGURE 4.9: Résultats de la trajectoire optimale pour une source sonore à 1.20 m et 90°

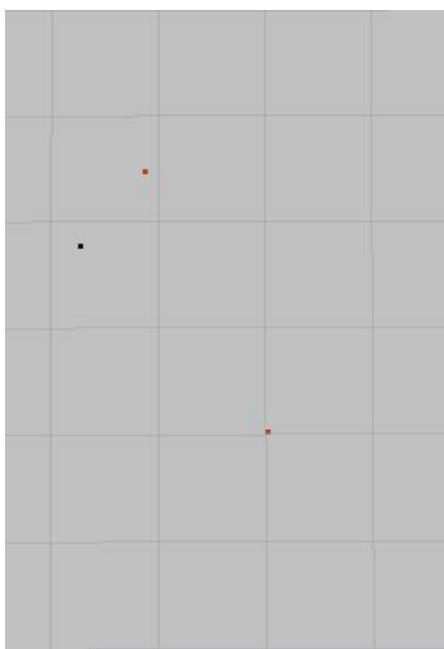


(A) Positions de départ et finale du ro-
bot, position de la source

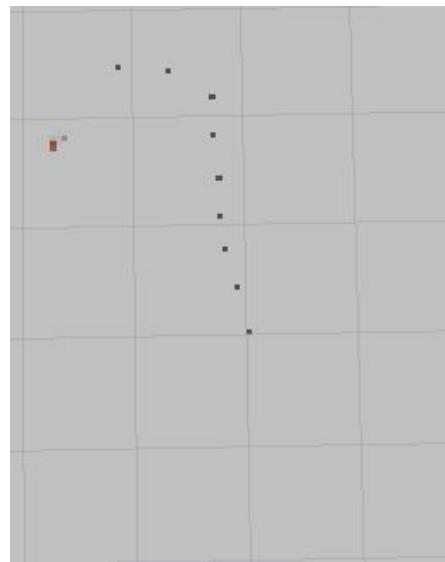


(B) Position de la source estimée et tra-
jectoire du robot mobile

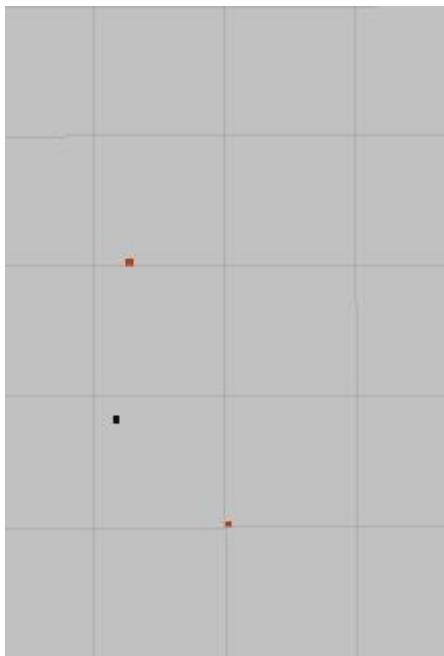
FIGURE 4.10: Résultats de la trajectoire optimale pour une source sonore à 2.40 m et 90°



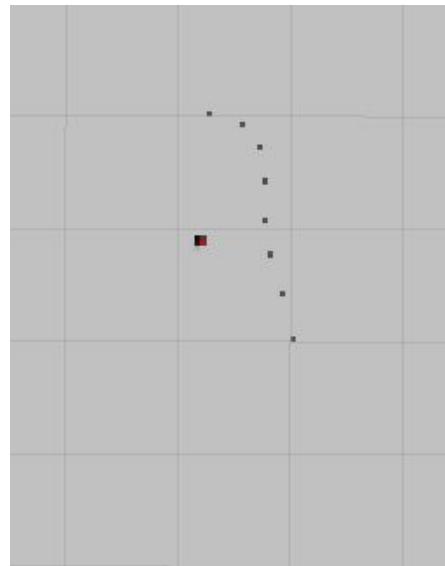
(A) Positions de départ et finale du robot, position de la source



(B) Position de la source estimée et trajectoire du robot mobile

FIGURE 4.11: Résultats de la trajectoire optimale pour une source sonore à 2.40 m et 135° 

(A) Positions de départ et finale du robot, position de la source



(B) Position de la source estimée et trajectoire du robot mobile

FIGURE 4.12: Résultats de la trajectoire optimale pour une source sonore à 1.20 m et 135°

Les résultats présentés dans les tableaux suivants :

nombres de positions	procédure automatisée (m)	procédure aléatoire (m)
2	0,33	0,54
4	0,33	0,43
6	0,16	0,32
8	0,09	0,23

TABLE 4.2: résultat de localisation de la source sonore placée à 45° et à 1.2 m de la position de départ du robot

nombre de positions	procédure automatisée (m)	procédure aléatoire(m)
2	0,36	0,49
4	0,26	0,40
6	0,16	0,29
8	0,08	0,19

TABLE 4.3: résultat de localisation de la source sonore placée à 135° et à 1.2 m de la position de départ du robot

nombres de positions	procédure automatisée (m)	procédure aléatoire (m)
2	0,78	0,99
4	0,54	0,83
6	0,37	0,64
8	0,2	0,46
10	0,13	0,38
12	0,08	0,3

TABLE 4.4: résultat de localisation de la source sonore placée à 135° et à 2.4 m de la position de départ du robot

nombres de positions	procédure automatisée (m)	procédure aléatoire (m)
2	0,76	1,0
4	0,50	0,88
6	0,35	0,73
8	0,18	0,60
10	0,13	0,44
12	0,08	0,34

TABLE 4.5: Résultat de localisation pour une source sonore placée à 45° et 2.4 m par rapport au repéré fixe

nombres de positions	procédure automatisée (m)	procédure aléatoire (m)
2	0,60	0,75
4	0,43	0,60
6	0,23	0,49
8	0,18	0,38
10	0,14	0,25
12	0,05	0,23

TABLE 4.6: Résultat de localisation pour une source sonore placée à 90° et à 2.4 m de la position de départ du robot

nombres de positions	procédure automatisée (m)	procédure aléatoire (m)
2	0,29	0,32
4	0,17	0,25
6	0,08	0,20
8	0,05	0,15

TABLE 4.7: Résultat de localisation pour une source sonore placée à 90° et à 1.2 m de la position de départ du robot

Chapitre 5

Conclusion et perspectives

5.1 Conclusion

Dans le cadre de ce travail, nous avons amélioré les capacités du robot mobile à localiser une source sonore et à construire une carte de la scène auditive. Le robot se déplace dans une grille d'occupation, qui met à jour à chaque pas, à chaque mesure enregistrée par la Kinect et traitée par l'algorithme *MUSIC – GSVD*. Pour chaque mesure et à chaque pas, le robot planifie une trajectoire optimale qui lui permettra de localiser la source d'une façon précise. L'interprétation et l'intégration des informations données par les mesures combinées au déplacement du robot permettent de construire une carte de l'environnement sonore.

Cette même procédure permet de guider le robot mobile dans ses tâches, et équipe de cette façon le robot d'une technologie, qui va lui faciliter la prise de décision, et le rendre plus efficace lors des missions de navigation, de surveillance ou d'exploration des terrains inconnus.

5.2 Perspectives

Dans ce travail, on a considéré que la source sonore est à la même hauteur que les capteurs. Nous avons comme perspective de construire un modèle de l'environnement en 3D, qui nous permettra l'étude de la scène auditive en 3D et la cartographie de l'environnement. Le déplacement du robot mobile et la position des sources peuvent être modélisés par des *Processus de Markov Partiellement observable (POMDP)* ou ρ *POMDP* basé sur la croyance qu'on a sur la position des sources sonores et du robot mobile.

Une fois que les sources sonores sont localiser, il est possible de considérer la dynamique des sources et leurs orientations dans l'environnement pour ensuite faire du suivi. À partir des données de localisation, on peut faire de la séparation de sources et de la reconnaissance de la parole.

Bibliographie

- [1] Ui-Hyun Kim and Hiroshi G.Okuno. Improved binaural sound localization and tracking for unknown time-varying number of speakers. *Advanced Robotics*, pages 1161–1173.
- [2] Jean Rouat François Michaud Kazuhiro Nakadai Hiroshi G.Okuno Jean-Marc Valin, Shun’ichi Yamamoto. Robust recognition of simultaneous speech by a mobile robot. *Robotics*, 23 :742–752, 2007.
- [3] Futoshi Asano Masataka Goto, Katunobu Itou and Hideki Asohsss. Real-time sound source localization and separation system and its application to automatic speech recognition. *International Conference Eurospeech.*, pages 1013–1016, 2001.
- [4] Matsui T. Asono F., Asoh H. Sound source localization and signal separation for office robot “jijo-2”. *Multisensor Fusion and Integration for Intelligent Systems. MFI ’99. Proceedings.IEEE/SICE/RSJ International Conference*, page 243 – 248, 1999.
- [5] Keisuke Nakamuras Kazuhiro Nakadai and Gökhan Ince. Real-time super-resolution sound source localization for robots. *IEEE, International Conference on Intelligent Robots and Systems*, pages 694–699, 2012.
- [6] Charles Blandin Alexey Ozerov and Emmanuel Vincent. Multi-source tdoa estimation in reverberant audio using angular spectra and clustering. *Preprint submitted to Signal Processing*, 92 :1950–1960, 2012.
- [7] Ralph 0.Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 3 :276–280, 1986.
- [8] Dominic Letourneau Jean Rouat Jean Marc-Valin, François Michaud. Robust sound source localization using a microphone array on a mobile robot. *Robotics and Automation, 2004. Proceedings. ICRA ’04. IEEE International*, 1 :1033 – 1038, 2004.
- [9] Nehorai Arye Stoica Petre. Music, maximum likelihood, and cramer-rao bound. *Acoustics, Speech and Signal Processing, IEEE*, 37 :720–741, 2002.

- [10] Sebastian Thrun. Learning occupancy grid maps with forward sensor models. *Autonomous Robots*, pages 111–127, 2003.
- [11] Alberto Elfes. Using occupancy grids for mobile robot perception and navigation. *IEEE ASSP Magazine*, 22 :45–57, 1989.
- [12] Alberto Elfes. Occupancy grids a stochastic spatial representation for active robot perception. *IEEE ASSP Magazine*, 22 :136–146, 1989.
- [13] Alberto Elfes. Dynamic control of robot perception using stochastic spatial models. *International Workshop on Information Processing in Mobile Robots*, 3 :2561–2567, 1992.
- [14] Edmond Boyer Jean-Sébastien Franco. Fusion multi-vue d’informations de silhouettes à l’aide d’une grille d’occupation 3d. *15e congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle 2*, pages 1747–1753, 2006.
- [15] E.Martinson and A.Schultz. Robotic discovery of the auditory scene. *Robotics and Automation*, pages 435–440, 2007.