

ED SMAER

Sujet de thèses 2013

Laboratoire : ISIR (Institut des Systèmes Intelligents et de Robotique)

Etablissement de rattachement : Université Pierre et Marie Curie

Directeur de thèse et section CNU ou CNRS : Bruno GAS, 61^{ème} section

Codirection et section CNU et CNRS : Sylvain ARGENTIERI, 61^{ème} section

Titre de la thèse : **Binaural Active Exploratory Listening in Robotics**

Collaborations dans le cadre de la thèse : LAAS-CNRS (Toulouse, France)

Rattachement à un programme : Projet européen déposé en avril 2013 (TWO!EARS), Consortium européen « *Aural Assessment By Means of Binaural Algorithms (AABBA)* »

Résumé du sujet

L'analyse computationnelle de scène auditive dans un cadre binaural reste aujourd'hui encore un défi. Peu robustes aux conditions sonores difficiles (incluant bruit, réverbérations, etc.), les méthodes actuelles n'exploitent que rarement une possibilité fondamentale du système auditif: le mouvement. Nous avons démontré récemment dans un contexte robotique comment ce mouvement, combiné aux variations perceptives auditives, étaient susceptibles d'améliorer significativement l'analyse de la scène sonore. Cette thèse propose de continuer dans cette voie, avec pour objectif d'étudier l'apport du mouvement pour la localisation, mais également pour des tâches auditives plus haut niveau (telles que la reconnaissance du locuteur), et éventuellement multimodales.

Sujet développé

La thématique de l'audition binaural vise à analyser une scène sonore à l'aide des signaux perçus par une paire de microphones, généralement placés sur une tête anthropomorphe. Sur la base de ces signaux, l'objectif est alors d'être capable de localiser, extraire, reconnaître et interpréter les différents sons présents autour du capteur binaural. Seulement, dans les environnements difficiles et dynamiques de la Robotique --où bruits et réverbérations se combinent avec les signaux d'intérêts--, les résultats de cette analyse restent limités. Dans ce cadre, la localisation de source sonore est fondamentale : ce n'est généralement qu'après avoir localisé précisément un son dans l'environnement du robot que l'on peut envisager la mise en place de méthodes d'extraction et reconnaissance de la source d'intérêt. Sans cette étape, il est illusoire d'espérer doter un robot de capacités de dialogue avec un humain si ce dernier n'est pas équipé d'un casque avec microphone permettant l'acquisition de signaux de parole dans de bonnes conditions. De fait, l'analyse computationnelle de scène auditive dans un cadre binaural reste donc aujourd'hui encore un défi.

Cependant, peu robustes aux conditions sonores difficiles, les méthodes et algorithmes actuels n'exploitent que rarement une possibilité fondamentale du système auditif : le mouvement. Des stratégies très simples et naïves peuvent permettre en effet d'améliorer sensiblement l'analyse de la scène sonore : tourner la tête pour lever l'ambiguïté devant/derrière en localisation, se rapprocher du locuteur afin d'améliorer les résultats de la reconnaissance de parole, etc. Nous avons démontré récemment --au cours d'une thèse co-encadrées entre le LAAS (Toulouse, France) et l'ISIR-- comment ce mouvement, combiné aux variations perceptives auditives résultantes, était susceptible d'améliorer significativement l'analyse de la scène sonore. On parle alors d'audition binaurale *active*. Les travaux menés ont principalement porté sur la formalisation précise du problème de la localisation d'une source (possiblement en mouvement) par un capteur binaural à l'aide d'algorithmes de filtrage stochastique. L'idée consiste ici à fusionner la proprioception du robot avec les variations d'indices sonores extraits des signaux binauraux afin d'estimer dans un cadre probabiliste la position de la source d'intérêt. L'essentiel des travaux a été principalement validé en simulation et en environnements maîtrisés (sans bruit ni réverbération) en présence d'une seule et unique source sonore.

Cette thèse propose de continuer dans cette voie, avec pour objectif de valider expérimentalement l'apport du mouvement pour la localisation, mais également pour des tâches auditives plus haut niveau (telles que la reconnaissance du locuteur), possiblement multimodales. Plus précisément, les travaux porteront essentiellement sur :

- la constitution et la maintenance d'une base de donnée contenant différentes scènes audio-visuelles. Cette base de donnée sera constituée de signaux binauraux, de séquences vidéos, de mesures de réalité terrain (obtenue à l'aide d'un système de capture de mouvement), et précisera les conditions acoustiques d'enregistrement (taille de la pièce, présence d'obstacles, temps de réverbération, etc.) Les données recueillies devront être annotées précisément, et respecter certains scénarii définis auparavant. Les enregistrements auront lieu dans la salle son de l'ISIR (salle quasi-anéchoïque), mais également dans des environnements plus réalistes à difficulté acoustique croissante ;
- la validation expérimentale des algorithmes de localisation active proposés par le groupe. Au delà des aspects théoriques et fondamentaux, une attention particulière sera portée à l'implémentation matérielle et logicielles des méthodes. Une plateforme robotique dédiée sera utilisée à cet effet, couplée à une implémentation sous ROS de modules de traitement audio (allant du bas niveau --extractions de caractéristiques audio, calculs de « cartes binaurales »-- au haut niveau --tracking, décision--.)

Le candidat devra avoir des connaissances en traitement du signal et filtrage, mais également de fortes compétences en programmation (C++, Matlab).