

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي و البحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

المديرية العامة للبحث العلمي و التطوير التكنولوجي

Direction Générale de la Recherche Scientifique et du Développement Technologique



## Projet national de recherche : rapport final

# SYSTÈME DE SUPERVISION DE CAMÉRA BASÉ SUR L'INGÉNIERIE AUDIO

**Chef de projet : SAYOUD Halim**

**Affiliation : Faculté d'électronique et d'informatique USTHB**

**Université : USTHB**

**Organisme pilote : CERIST**

**Programme national de Recherche : 2011-2013**



وزارة التعليم العالي و البحث العلمي  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
المديرية العامة للبحث العلمي و التطوير التكنولوجي  
Direction Générale de la Recherche Scientifique et du Développement Technologique

**تقرير عام لمشروع البحث**  
**Rapport général du projet PNR**

# Rapport général du projet PNR 2013

**Titre du projet :**

12/u160/5155

**CERIST**

## **Système de Supervision de Caméra basé sur l'Ingénierie Audio**

**Chef du projet : Prof SAYOUD Halim**

**Université USTHB**

Nom et prénom الاسم و اللقب	Grade الرتبة	Etablissement employeur المؤسسة المستخدمة	Observation
SAYOUD Halim	Prof	USTHB	Membre Chef du projet
HOUACINE Amrane	Prof	USTHB	Membre
MEKAOUI Slimane	MCA	USTHB	Membre
OUAMOUR Siham	MCA	USTHB	Membre
KHENNOUF Salah	MAA	U. Msila	Membre

Octobre 2013

تقرير عام لمشروع البحث  
Rapport général du projet PNR

I-Identification du projet:

PNR

12/u160/5155

1-التعريف بالمشروع

Organisme pilote

CERIST

Domiciliation du projet :

Université USTHB  
Faculté d'électronique et d'Informatique  
Labo LCPTS  
USTHB

Intitulé du projet

عنوان المشروع

Système de Supervision de Caméra basé sur l'Ingénierie Audio

Chercheurs impliqués dans le projet

أعضاء المشروع و المؤسسة المستخدمة

Nom et prénom الاسم و اللقب	Grade الرتبة	Etablissement employeur المؤسسة المستخدمة	Observation
SAYOUD Halim	Prof	USTHB	Membre Chef du projet
HOUACINE Amrane	Prof	USTHB	Membre
MEKAOUI Slimane	MCA	USTHB	Membre
OUAMOUR Siham	MCA	USTHB	Membre
KHENNOUF Salah	MAA	U. Msila	Membre

Déroulement du projet :

Le travail de recherche que nous proposons de réaliser émane de plusieurs disciplines : le traitement du signal audio ; la robotique et l'électronique. Il s'intéresse à la supervision des enregistrements audio-visuels par caméra dans des meeting-rooms. De tels studios sont appelés, parfois, des smart-rooms (en Français : salle (ou studio) intelligente).

Ce travail de recherche se voit destiné à la supervision des multi-conférences et séminaires, utilisant l'information audio pour la localisation et le suivi du locuteur parlant.

En effet, nous tentons de réaliser et commander un système automatique pour orienter une ou plusieurs caméras mobiles ou fixes (à focaliser) vers des cibles sonores, représentées dans notre cas

par des locuteurs fixes ou mobiles.

Les principaux objectifs de ce travail sont :

- d'une part, le développement et l'implémentation d'algorithmes d'identification, de localisation et de suivi des locuteurs actifs par des techniques d'ingénierie audio et la mise au point d'un système automatique de supervision, permettant le suivi du locuteur parlant (*locuteur actif*) par une caméra mobile ;
- et d'autre part, la participation à la formation de doctorants et thésards dans ce nouveau domaine de recherche.

Les étapes et taches du projet sont :

- 1- Mise à niveau et Formation ;
- 2- Acquisition de matériel spécifique ;
- 3- Enregistrement des bases de données audio-visuelles ;
- 4- Implantation des méthodes de détection et localisation sonore ;
- 5- Implantation des méthodes de contrôle de caméra ;
- 6- Réalisation de cartes électroniques et Expérimentation ;
- 7- Publications et Validation de notre système (et des résultats).

# Table des matières

<b>Rapport d'activité</b>	1
<b>Rapport scientifique</b>	21
Introduction	24
Contenu du travail (théorie et Expérimentation)	26
Chapitre I GENERALITES SUR LA LOCALISATION DE SOURCE SONORE	26
Chapitre II BASE DE DONNEES REALISEE	37
Chapitre III METHODES DE LOCALISATION SONORE PROPOSEES	45
Chapitre IV SYSTEMES PROPOSES POUR LA COMMANDE DE CAMÉRA	54
Chapitre V RESULTATS	58
Chapite VI Publications associées au projet	65
Conclusions	67
Bibliographie	69
Annexes	74
Information financière	74

# RAPPORT D'ACTIVITÉ

**Titre du projet :**

12/u160/5155

**CERIST**

## **Système de Supervision de Caméra basé sur l'Ingénierie Audio**

Nom et prénom الاسم و اللقب	Grade الرتبة	Etablissement employeur المؤسسة المستخدمة	Observation
SAYOUD Halim	Prof	USTHB	Membre Chef du projet
HOUACINE Amrane	Prof	USTHB	Membre
MEKAOUI Slimane	MCA	USTHB	Membre
OUAMOUR Siham	MCA	USTHB	Membre
KHENNOUF Salah	MAA	U. Msila	Membre

## RAPPEL DES OBJECTIFS DU PROJET FIXES INITIALEMENT

- 1- Mise à niveau et Formation ;
- 2- Acquisition de matériel spécifique ;
- 3- Enregistrement des bases de données audio-visuelles ;
- 4- Implantation des méthodes de détection et localisation sonore ;
- 5- Implantation des méthodes de contrôle de caméra ;
- 6- Réalisation de cartes électroniques et Expérimentation ;
- 7- Publications et Validation de notre système (et des résultats).

## TACHES PREVUES ET REALISEES

Tâches prévues dans le projet	Tâches effectivement réalisées
1- Mise à niveau et Formation ;	Tache réalisée
2- Acquisition de matériel spécifique ;	Tache réalisée
3- Enregistrement des bases de données audio-visuelles ;	Tache réalisée
4- Implantation des méthodes de détection et localisation sonore ;	Tache réalisée
5- Implantation des méthodes de contrôle de caméra ;	Tache entamée mais non achevée à 100%
6- Réalisation de cartes électroniques et Expérimentation ;	Tache entamée mais non achevée à 100%
7- Publications et Validation de notre système et des résultats	Tache réalisée Voir le lien suivant de la vidéo : <a href="http://sayoud.net/images/PNR/New%20Video%20SAM_3605.MPG">http://sayoud.net/images/PNR/New%20Video%20SAM_3605.MPG</a>

## Résultats obtenus

### **-CHAPITRES DE LIVRE EDITES EN 2012/2013 :**

Nom du Livre : Advancing the Next-Generation of Mobile Computing: Emerging Technologies.

Titre: Automatic Speaker Localization and Tracking: Using a Fusion of the Filtered Correlation with the Energy Differential.

Site Web : <http://www.igi-global.com/book/advancing-next-generation-mobile-computing/58266>

### **PUBLICATIONS DANS DES REVUES INTERNATIONALES : 3 revues internationales**

Lien: <http://sayoud.net/Publications.html>

#### **1ere Revue internationale**

Journal: International Journal of Speech Technology - Springer.

Title: A New Approach of Speaker Clustering based on the Stereophonic Differential Energy.

Authors: S. Ouamour, H. Sayoud

References: Springer, Int. Jo. of Speech Techn. 2013; DOI: 10.1007/s10772-013-9199-z.

Website: <http://www.springer.com/engineering/signals/journal/10772>

#### **2ème Revue internationale**

Journal: International Journal of Speech Technology

Publisher: Springer Verlag

Indexed by: INSPEC (IET)

Title : A pertinent learning machine input feature for speaker discrimination by voice

Received: 31 August 2011 / Accepted: 23 January 2012. pp 181-190. Vol 15, no 02.

© Springer Science+Business Media, LLC 2012

Authors: S. Ouamour · H. Sayoud



References: Received: 31 August 2011 / Accepted: 23 January 2012

Int J Speech Technol (2012) 15:181–190

Website : <http://www.springer.com/engineering/signals/journal/10772>

### **3ème Revue internationale**

Journal: International Journal of Networking and Virtual Organisations,

Publisher: Indersciences, Switzerland

Indexed by: INSPEC (IET)

Title : Virtual Speaker Tracking by Camera Using a Sound Source Localization with two Microphones

Authors: H. Sayoud, S. Khenouf and S. Ouamour

References: Vol. 12, No. 2, 2013. pp: 85-110.

Website : <http://www.inderscience.com/browse/index.php?journalID=66&action=coming>

Copyright © 2013 Inderscience Enterprises Ltd.

### **CONFERENCES INTERNATIONALES**

Lien : <http://sayoud.net/publications.html>

#### **1- ICTA-2013 conference – Hammamet, Tunisia, 24-26 October, 2013.**

Automatic Speaker Localization based on Speaker Identification -A Smart Room Application-

<http://www.icta.rnu.tn/index.php?picta=1>

#### **2- STA-2012 conference - Monastir, Tunisia, December 17-19, 2012.**

Automatic Speaker Detection Using Support Vector Machines and an Interlaced Segmentation - Application on Telephonic Talks-. Accepted in the 11th International conference on Sciences and Techniques of Automatic control & computer engineering. Authors: Sayoud et al. <http://www.sta-tn.com> , pp 1517-1525.

**3- ICCSE 2012 - London, UK, July 4-6, 2012,**

Virtual System of Speaker Tracking by Camera Using an Audio-Based Source Localization.  
Conference ICCSE, July 4-6, 2012, London, UK. pp 819-822.

**4- WCE 2011 conference - London, U.K., 6-8 July, 2011.**

SPEAKER CLUSTERING OF STEREOPHONIC SPEECH SIGNAL USING  
SPATIAL AND SEQUENTIAL GATHERING. World Congress on Engineering  
2011. Pp 1206-1210

**5- WOSSPA 2011 conference - Tipaza May 9-11 2011.**

Speaker localization using stereo-based sound source localization. In Systems, Signal Processing  
and their Applications (WOSSPA), 2011 7th International Workshop on (may 2011), pp. 231-234.

**-BASES DE DONNEES AUDIOVISUELLES FABRIQUEES (SUR DVD) EN 2013 :**

Nous avons fabriqué une grande base de données audio contenant plusieurs scénarios et plusieurs  
locuteurs. Nous l'avons appelé ASDB2, les détails peuvent être retrouvés sur le site web :

<http://sayoud.net/ASDB-2.pdf>

**-ENCADREMENT D'ETUDIANTS EN PFE :**

Plusieurs étudiants de Master ont été encadrés et formés en 2012 et 2013 : 3 thèses de Master ont  
été soutenues (voir les détails plus bas).

**Contraintes Rencontrées :**

- Matériel livré tardivement ;
- Partenaire économique (technicien) souvent non disponible ;
- Certains instruments de pointe font défaut (exemple : microphones matriciels).

## Objectifs

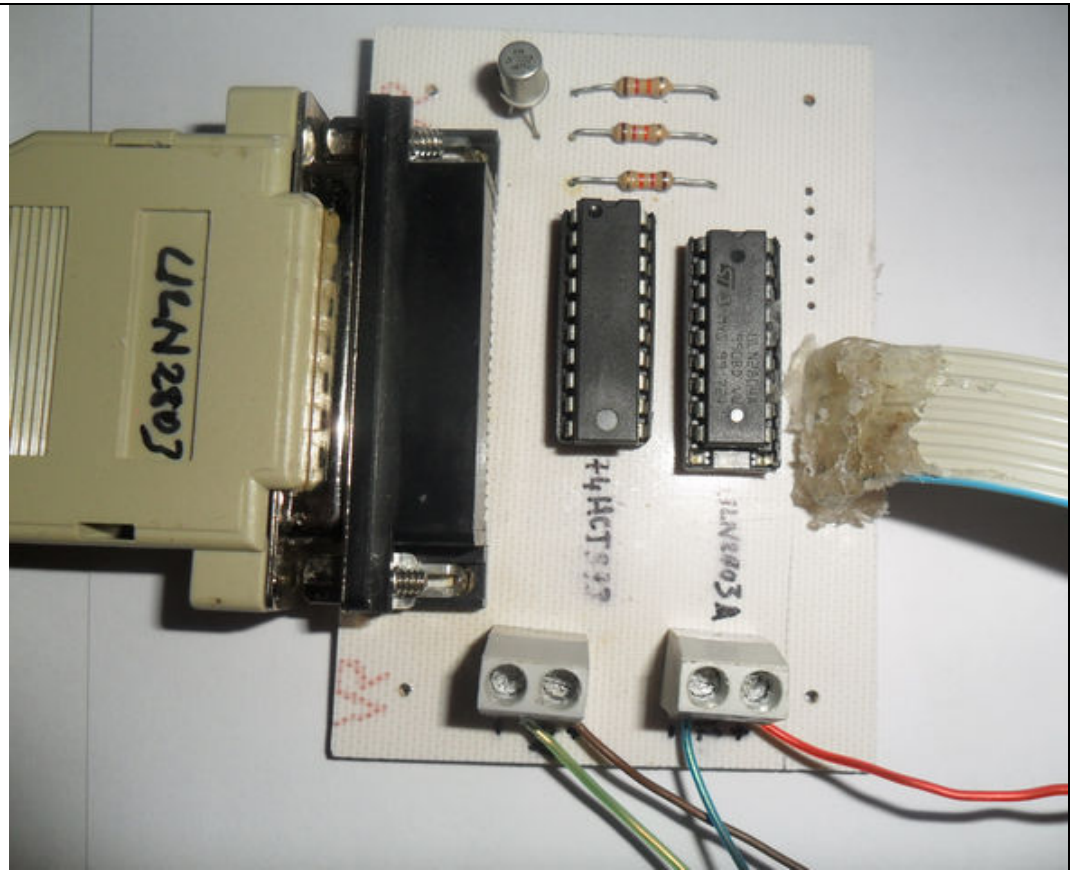
critère	Description
Conformité	<p>Les résultats obtenus sont assez conformes à ce qui devrait être obtenu théoriquement, mais on ne peut pas dire que c'est terminé à 100%. On a trouvé quelques problèmes de finalisation, tels que la partie commande et le fonctionnement en temps réel.</p> <p>Voir le lien suivant de la vidéo :</p> <p><a href="http://sayoud.net/images/PNR/New%20Video%20SAM_3605.MPG">http://sayoud.net/images/PNR/New%20Video%20SAM_3605.MPG</a></p>
Innovation	<p>Plusieurs innovations ont été apportées, en l'occurrence :</p> <ul style="list-style-type: none"> <li>-de nouvelles méthodes pour la détection de source audio (publiée dans des revues internationales – <i>voir ci-haut</i>) ;</li> <li>-des méthodes de commande virtuelle ont été inventées et publiées dans des revues internationales – <i>voir ci-haut</i>.</li> </ul>
Taux de réalisation	<p>Le taux de réalisation est environ estimé à moins de 80%.</p> <p>Une partie temps-réel aurait été souhaitée pour finaliser le projet.</p>
Méthodologie	<p>La méthodologie entreprise était bien étudiée :</p> <ul style="list-style-type: none"> <li>-Recherche bibliographique ;</li> <li>-mise à niveau ;</li> <li>-Implantation d'algorithmes appropriés et essai d'inventer ce qui n'existe pas encore dans ce domaine de recherche ;</li> <li>-Fabrication d'une base de données audio qu'on a appelée ASDB2</li> </ul> <p><a href="http://sayoud.net/ASDB-2.pdf">http://sayoud.net/ASDB-2.pdf</a>, pour les scénarios audiovisuels ;</p>

	<p>-Expériences de tests sur la Base de données DB2 (conçue auparavant) ;</p> <p>-Réglage, après mis au point, de cartes électroniques et mécatroniques avec tests expérimentaux ;</p> <p>-Publication, après exploitation, des résultats obtenus dans des revues internationales et des actes de proceeding de conférences ;</p> <p>-Mise au point de logiciels utiles pour la communauté de l'audiovisuel.</p>
--	--

## Résultats

Critère	Description
Brevets d'invention	<p>Nous avons des prototypes intéressants mais qui ne sont pas encore brevetés.</p> <p>Voir le lien suivant de la vidéo :</p> <p><a href="http://sayoud.net/images/PNR/New%20Video%20SAM_3605.MPG">http://sayoud.net/images/PNR/New%20Video%20SAM_3605.MPG</a></p>
Brevets d'innovation	<p>Idem : nous avons des prototypes intéressants mais qui ne sont pas encore brevetés</p>
Implication et satisfaction du partenaire socio économique (attestation ou convention)	<p>Nous résumons cet aspect relatif au partenaire par les points suivants :</p> <p>-Encadrement d'une enseignante du partenaire économique (ISMAS) en Doctorat : Mlle Kennai Wassila, enseignante chez le partenaire économique, qui s'est inscrite en Doctorat avec Mme Ouamour Siham (Membre du projet) ;</p> <p>-Proposition de coopération pour l'ouverture d'une nouvelle filière de Licence chez le partenaire économique (ISMAS) : Prof Houacine (membre du projet) s'est engagé à leur assurer des formations ;</p> <p>-Discussions et échange d'idées intéressants entre notre équipe et le technicien du partenaire économique. Les idées jaillissant de ces discussions ont été assez intéressantes.</p>

Prototype	<p>-Deux prototypes de cartes électroniques (1 en maquette et 1 en circuit imprimé) ;</p> <p>-Deux systèmes mécatroniques de commande du moteur pas-à-pas (voir ci-dessous) ;</p>  <p><i>Photo du système électronique et mécatronique</i></p>



*Photo d'une des cartes électroniques*

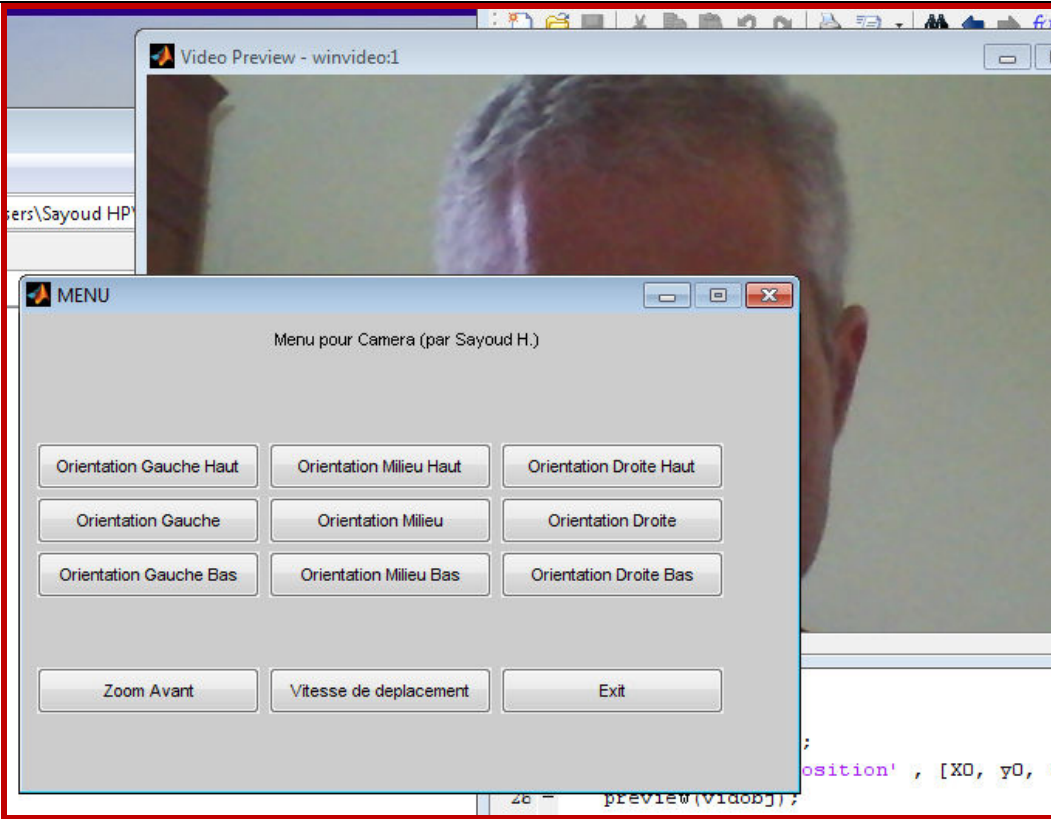


*Photo du système électronique, mécatronique et de la caméra*

#### Logiciels

Les softwares suivants ont été élaborés :

- Logiciel de détection de position basé sur la FCM, EDM et la GCC-PHAT;
- Logiciel de rotation de la caméra utilisant le port parallèle ;
- Logiciel de commande virtuelle, dont la version exécutable est à la disposition de la communauté scientifique, économique et industrielle, sur simple demande. Voir ci-dessous :

	 <p style="text-align: center;"><i>Photo du logiciel de commande virtuelle de la caméra</i></p>
Coopération internationale	<p>-Stage de Mr Khennouf Salah (membre du projet) en Malaisie</p> <p>-Présentation (sous forme de Séminaire) concernant le sujet du projet suivi d'une discussion à l'université internationale " IIUM " International Islamic University Malaysia" lors d'un Meeting Mensuel de l'IMSRU " Intelligent Mechatronics System Research Unit" en Malaisie. <a href="http://www.iium.edu.my/imsr/">http://www.iium.edu.my/imsr/</a></p>
Publications (nationales, internationales)	<p><b>PUBLICATIONS DANS DES REVUES INTERNATIONALES : 3 revues internationales dans le thème du projet</b></p> <p>Lien: <a href="http://sayoud.net/publications.html">http://sayoud.net/publications.html</a></p> <p><b>1er journal</b></p> <p>Journal: International Journal of Speech Technology - Springer.</p> <p>Title: A New Approach of Speaker Clustering based on the Stereophonic Differential Energy.</p>



Authors: S. Ouamour, H. Sayoud

References: Springer, Int. Jo. of Speech Techn. 2013; DOI: 10.1007/s10772-013-9199-z.

Website: <http://www.springer.com/engineering/signals/journal/10772>

### **2ème journal**

Journal: International Journal of Speech Technology

Publisher: Springer Verlag

Indexed by: INSPEC (IET)

Title : A pertinent learning machine input feature for speaker discrimination by voice

Accepted: 23 January 2012. pp 181-190. Vol 15, no 02.

© Springer Science+Business Media, LLC 2012

Authors: S. Ouamour · H. Sayoud

Int Journal of Speech Technology (2012) 15:181–190

Website : <http://www.springer.com/engineering/signals/journal/10772>

### **3ème journal**

Journal: International Journal of Networking and Virtual Organisations,

Publisher: Indersciences, Switzerland

Indexed by: INSPEC (IET)

Title : Virtual Speaker Tracking by Camera Using a Sound Source Localization with two Microphones

Authors: H. Sayoud, S. Khennouf and S. Ouamour

References: Vol. 12, No. 2, 2013. pp: 85-110.

Website :

<http://www.inderscience.com/browse/index.php?journalID=66&action=coming>

Copyright © 2013 Inderscience Enterprises Ltd.

<p>Communications (nationales, internationales)</p>	<p style="text-align: center;"><b>CONFERENCES INTERNATIONALES</b></p> <p><b>Lien:</b> <a href="http://sayoud.net/publications.html">http://sayoud.net/publications.html</a></p> <p><b>1- ICTA-2013 conference – Hammamet, Tunisia, 24-26 October, 2013.</b> Automatic Speaker Localization based on Speaker Identification -A Smart Room Application- Tunisia, 24-26 October, 2013. <a href="http://www.icta.rnu.tn/index.php?picta=1">http://www.icta.rnu.tn/index.php?picta=1</a></p> <p><b>2- STA-2012 conference - Monastir, Tunisia, December 17-19, 2012.</b> Automatic Speaker Detection Using Support Vector Machines and an Interlaced Segmentation -Application on Telephonic Talks-. In the 11th International conference on Sciences and Techniques of Automatic control &amp; computer engineering. Authors: Sayoud et al. <a href="http://www.sta-tn.com">http://www.sta-tn.com</a> , pp 1517-1525.</p> <p><b>3- ICCSE 2012 conference - London, UK, July 4-6, 2012,</b> Virtual System of Speaker Tracking by Camera Using an Audio-Based Source Localization. Conference ICCSE, July 4-6, 2012, London, UK. pp 819-822.</p> <p><b>4- WCE 2011 conference - London, U.K., 6-8 July, 2011.</b> SPEAKER CLUSTERING OF STEREOPHOIC SPEECH SIGNAL USING SPATIAL AND SEQUENTIAL GATHERING. World Congress on Engineering 2011. Pp 1206-1210</p> <p><b>5- WOSSPA 2011 conference - Tipaza May 9-11 2011.</b> Speaker localization using stereo-based sound source localization. In Systems, Signal Processing and their Applications (WOSSPA), 2011 7th International Workshop on (may 2011), pp. 231-234.</p>

Encadrement avec soutenance de master, magister ou doctorat	<p>-Encadrement d'un Master (soutenu en 2013) dans le thème du projet (Université de Msila). Titre : Réalisation d'un Système Automatique pour la Commande d'une Camera vers des Locuteurs Actifs. Celui-ci a soutenu en 2013. Encadreurs : H. Sayoud et S. Khennouf. Thésard : Aiche abderrazak.</p> <p>-Encadrement de 2 étudiantes en Master (Master soutenu en 2012) dans le thème du projet (Université de Medea). Titre : Détection des Locuteurs par des Signaux Stéréophoniques. Celles-ci ont soutenu en 2012. Encadreurs : S. Ouamour et H. Sayoud. Thésards : Djemai Lynda et Sadi Hadjer.</p> <p>- Encadrement d'un étudiant en Master. Titre : Suivi de séquence vidéo à l'aide de contours actifs, FEI, USTHB, Juin 2011. Encadreur : Prof A. Houacine.</p> <p>-Encadrement d'une enseignante du partenaire économique (ISMAS) en Doctorat (Mlle Kennai Wassila). Encadreurs : S. Ouamour et M. Guerti.</p>
---	---

## Impact socio économique

<b>Critère</b>	<b>Description</b>
Formation et perfectionnement de la ressource humaine	<p>-Enseignement d'un module de Master sur l'Acquisition Audio en étroit lien avec la première partie du projet. Enseignant : Prof H. Sayoud (Département télécommunication / Master TRM) ;</p> <p>-Encadrement de 3 étudiants en Master qui ont été formés durant ce projet (sur le thème du projet). Noms des étudiants : Aiche abderrazak, Djemai Lynda et Sadi Hadjer ;</p>

	<p>-Encadrement d'une enseignante du partenaire économique (ISMAS) en Doctorat. Nom de cette doctorante : Kennai Wassila ;</p> <p>-Stage de Mr Khennouf Salah (membre du projet) en Malaisie en 2013 ;</p> <p>-Présentation (sous forme de Séminaire) concernant le sujet de ce projet de recherche, suivi d'une discussion à " IIUM " International Islamic University Malaysia" lors d'un Meeting Mensuel de l'IMSRU " Intelligent Mechatronics System Research Unit", en septembre 2013 en Malaisie. Présentateur : Mr S. Khennouf (membre du projet).</p>
Création de nouvelles filières LMD	Proposition de coopération pour l'ouverture d'une nouvelle filière de Licence chez le partenaire économique (ISMAS). Professeur concerné : Pr A. Houacine (membre du projet).
Offre de service,	<p>-Mise a la disposition de la communauté scientifique d'un DVD contenant une base de données audio (ASDB2) <a href="http://sayoud.net/ASDB-2.pdf">http://sayoud.net/ASDB-2.pdf</a> pouvant être utile à des expériences scientifiques et didactiques ;</p> <p>-Des cartes électroniques et de nouveaux logiciels réalisés sont aussi disponibles. Lien : <a href="http://sayoud.net/images/PNR/New%20sam%203591.pdf">http://sayoud.net/images/PNR/New%20sam%203591.pdf</a></p> <p>-Logiciel de commande virtuel aidant la gestion de l'enregistrement vidéo par camera fixe, dont la version exécutable est à la disposition de la communauté scientifique, économique et industrielle, sur simple demande ;</p> <p>-Mise au point d'un cours didactique en parole (disponible sur le web) pour les techniciens de l'audio. Lien : <a href="http://sayoud.net/old/cours.pdf">http://sayoud.net/old/cours.pdf</a></p> <p>-Prototype du système mécatronique pour la commande de caméra.</p>
Adéquation du bilan avec le besoin socioéconomique initialement exprimé	<p>Plusieurs finalités ont été atteintes dans le cadre de ce projet (voir ci-dessus), bien que le projet soit assez vaste en tel.</p> <p>Le projet nous a été fructueux en recherche et en formation aussi bien.</p> <p>Aussi, nous aurions souhaité pouvoir assurer des séminaires que l'on a déjà proposés au partenaire économique. Nous n'avions pas eu le temps adéquat pour les faire.</p>

Réponse à la problématique posée par le partenaire socioéconomique	<p>La réponse à la problématique de réduction de cout et d'effort lors des enregistrements audiovisuels est claire. Cette automatisation assure en même temps la réduction du cout et de l'effort associés dans de grandes proportions. Elle permet aussi l'automatisation des studios d'enregistrement.</p> <p>Voir le lien de la vidéo :</p> <p><a href="http://sayoud.net/images/PNR/New%20Video%20SAM_3605.MPG">http://sayoud.net/images/PNR/New%20Video%20SAM_3605.MPG</a></p> <p>Par ailleurs, il en résulte les points suivants :</p> <ul style="list-style-type: none"> <li>- Diminution du surcout du tournage (<i>prises de vues audiovisuelles par caméra</i>) ;</li> <li>- Facilités apportées au metteur en scènes, lui permettant de diminuer l'effort physique et humain lors des prises de scènes ;</li> <li>- Innovation technologique dans le domaine Multimédia, audio et audiovisuel ;</li> </ul> <p>Néanmoins une mise en pratique sur le terrain nécessite encore quelques mises au point.</p>
--	--

## Valorisation

critère	Description
Pertinence du projet	<p>Le projet est assez pertinent vus les différents intérêts pratiques sur le terrain et les originalités de ce projet, dont on cite :</p> <ul style="list-style-type: none"> <li>-Ce projet touche un domaine très demandé qui est en pleine expansion ; il s'agit du domaine de l'audio-visuel.</li> <li>-Les besoins économiques dans ce domaine sont évidents et se vérifient par les grands chiffres d'affaire des compagnies de l'audio-visuel connues à l'échelle mondial.</li> <li>-Ce projet propose un système automatique de gestion de la camera par techniques exclusivement audio, d'où sa première originalité.</li> </ul>

	<p>Par ailleurs, pendant la phase préliminaire de production audiovisuelle, chaque étape est réellement établie pour concevoir précisément le film et de le planifier. En général, la société de production est créée et un office de production est mis en place. La production séquencée dans le story-board est visualisée avec l'aide des illustrateurs et des concepteurs. Un budget de production est établi pour planifier les dépenses du film. Pour les productions importantes, on a même des assurances qui sont prises pour fournir une protection contre les accidents éventuels.</p> <p>Le producteur audiovisuel embauche une grande équipe. La nature du film et son budget déterminent la taille de l'équipe utilisée durant la réalisation du tournage. Ainsi, de nombreux films à grand spectacle emploient des centaines d'acteurs et de personnels.</p> <p>Si un système automatique pouvait être mis au point, on aurait économisé toutes ses dépenses et emplois de personnels. A titre d'exemple, 1, 2 ou 3 techniciens devraient suffire largement pour l'opération d'enregistrement audio-visuel.</p> <p>Par ailleurs, les objectifs sont multiples :</p> <ul style="list-style-type: none"> <li>- Automatisation des studios d'enregistrement. Voir le lien de la vidéo : <a href="http://sayoud.net/images/PNR/New%20Video%20SAM_3605.MPG">http://sayoud.net/images/PNR/New%20Video%20SAM_3605.MPG</a> ;</li> <li>- Diminution du surcout du tournage (<i>prises de vues audiovisuelles par caméra</i>) ;</li> <li>- Facilités apportées au metteur en scènes, lui permettant de diminuer l'effort physique et humain lors des prises de scènes ;</li> <li>- Innovation technologique dans le domaine Multimédia, audio et audiovisuel ;</li> <li>- Intérêt didactique et aide à la formation des étudiants de l'audio-visuel du partenaire économique.</li> </ul>
<p>Caractère innovant du projet</p>	<p>Nous avons proposé un système intelligent, basé sur une seule caméra mobile à orientation automatique vers la cible sonore (1ere originalité).</p> <p>Voir le lien suivant de la vidéo : <a href="http://sayoud.net/images/PNR/New%20Video%20SAM_3605.MPG">http://sayoud.net/images/PNR/New%20Video%20SAM_3605.MPG</a></p>

Cela a fait l'objet de soutenances de thèses et de plusieurs publications dans des revues internationales (voir ci-haut).

Notre application est destinée à la supervision automatique des enregistrements audiovisuels dans les meeting-rooms. Plusieurs locuteurs assis autour d'une table, prennent la parole à tour de rôle et le système de contrôle des caméras suit le locuteur parlant en toute harmonie.

Pour ce faire nous avons proposé deux méthodes de localisation (*basées sur l'ingénierie audio*) : La première est basée sur la nouvelle technique de corrélation filtrée et la deuxième sur le différentiel d'énergie (2<sup>ème</sup> originalité).

Cela a fait l'objet de plusieurs publications internationales et conférences (voir ci-haut).

Pour la commande, on a proposé 2 méthodes: la première méthode est basée sur la commande par moteur pas à pas via un système automatique. La deuxième méthode est basée sur une nouvelle technique de commande virtuelle de caméra par le biais du contrôle de la ROI ou *Region Of Interest* (3<sup>ème</sup> originalité).

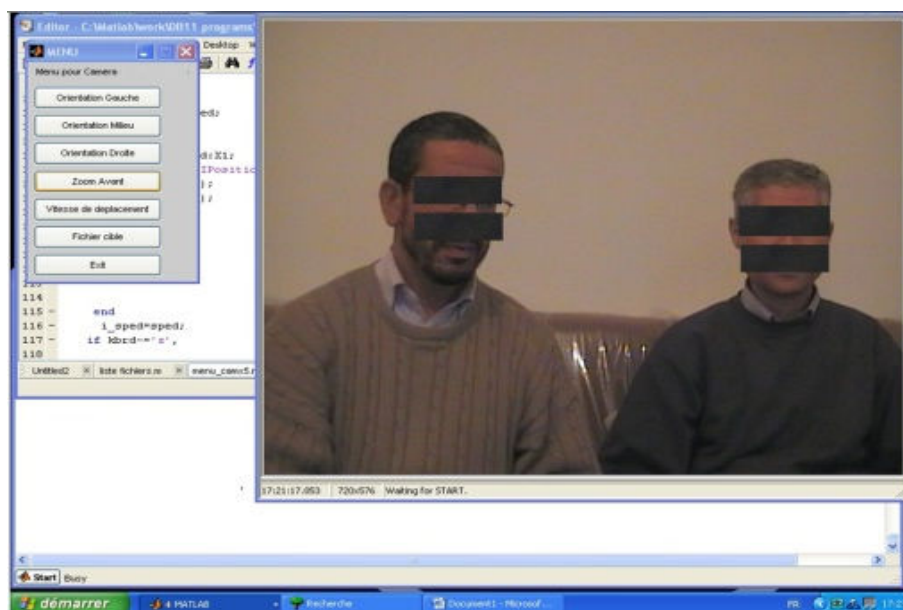
Nous avons appelé cette nouvelle méthode VROI ou *Virtual Region Of Interest technique*.

Par ailleurs, certaines cartes électroniques d'interfaçage et de commande ont déjà été réalisées auparavant. Celles-ci permettent le fonctionnement mécatronique du système (figure ci-dessous).



*Photo du système de commande, de la tourelle mécatronique et de la caméra.*

La nouvelle technique VROI est une technique majoritairement software qui permet de piloter les différentes régions capturées par la caméra virtuellement (voir figure suivante). Ceci est une première dans ce domaine de recherche (à notre connaissance).



*Photo du logiciel de VROI réalisé*



<p>Faisabilité du projet et son transfert vers le secteur socioéconomique</p>	<p>Ce projet comprend 2 grands types d'applications :</p> <ul style="list-style-type: none"> <li>-une application de formation, visant à former les thésards et doctorants dans ce nouveau domaine de recherche.</li> <li>-et une application industrielle qui consiste en l'automatisation de l'enregistrement audio-visuel ;</li> </ul> <p>Concernant la 1ère application, la faisabilité a été prouvée par des encadrements et des formations très spécialisées. Nous pensons que cette tâche reste aisée à réaliser au sein de notre université (laboratoires).</p> <p>Concernant la 2ème application, et pour pouvoir transférer ce projet vers le secteur socio-économique, il faut que nous proposons notre prototype à un partenaire industriel qui soit capable de le reproduire à un prix compétitif.</p>
<p>Impact du projet (amélioration de la qualité, réduction des coûts de production, réduction des délais, .....)</p>	<p>Au moyen terme, ce projet devrait conduire à plusieurs impacts intéressants, dont nous citons :</p> <ul style="list-style-type: none"> <li>-Formation de techniciens et ingénieurs dans le domaine de l'audio-visuel ;</li> <li>-Automatisation des studios d'enregistrement audio-visuel (smart meeting-rooms). Nous pouvons prévoir de tels systèmes à l'ENTV, par exemple ;</li> <li>-Cette dernière conduit à la réduction du coût de supervision dû au remplacement des caméramans par un système intelligent de pilotage, car au fait, chaque technicien nécessite un certain salaire lors d'un tournage AV ;</li> <li>-De même, il en résulte une fatigue moindre des techniciens de supervision (qui doivent être tout le temps attentifs), puisque ils seront assistés et parfois remplacés par cet automatisme intelligent ;</li> <li>-Un impact très important est que ce PNR a été une bouffée d'oxygène aux chercheurs dans le sens positif pour une avancée motivée en recherche appliquée ;</li> <li>-Une belle expérience poussant les laboratoires de recherche à sortir de la structure purement académique et de pouvoir écouter le secteur socio-économique en essayant de résoudre leurs éventuels problèmes ;</li> <li>-Amélioration des connaissances du personnel associé à ce projet en le rendant plus expérimenté et plus pratique sur le terrain socio-économique.</li> </ul>

# Table des matières du rapport scientifique

Résumé	22
Introduction	24
Contenu du travail (théorie et Expérimentation)	26
Chapitre I GENERALITES SUR LA LOCALISATION DE SOURCE SONORE	26
Chapitre II BASE DE DONNEES REALISEE	37
Chapitre III METHODES DE LOCALISATION SONORE PROPOSEES	45
Chapitre IV SYSTEMES PROPOSES POUR LA COMMANDE DE CAMÉRA	54
Chapitre V RESULTATS	58
Chapite VI Publications associées au projet	65
Conclusions	67
Bibliographie	69
Annexes	74
Information financière	74

# RESUME

## تقرير عام لمشروع البحث Rapport général du projet PNR

Direction Générale de la Recherche Scientifique et du Développement Technologique

### I-Identification du projet:

PNR

12/u160/5155

### 1-التعريف بالمشروع

Organisme pilote

CERIST

### Domiciliation du projet :

**Université USTHB**  
**Faculté d'électronique et d'Informatique**  
Labo LCPTS  
USTHB

### Intitulé du projet

عنوان المشروع

**Système de Supervision de Caméra basé sur l'Ingénierie Audio**

### Chercheurs impliqués dans le projet

أعضاء المشروع و المؤسسة المستخدمة

Nom et prénom الاسم و اللقب	Grade الرتبة	Etablissement employeur المؤسسة المستخدمة	Observation
SAYOUD Halim	Prof	USTHB	Membre Chef du projet
HOUACINE Amrane	Prof	USTHB	Membre
MEKAOUI Slimane	MCA	USTHB	Membre
OUAMOUR Siham	MCA	USTHB	Membre
KHENNOUF Salah	MAA	U. Msila	Membre

### Déroulement du projet :

Le travail de recherche que nous proposons de réaliser émane de plusieurs disciplines : le traitement du signal audio ; la robotique et l'électronique. Il s'intéresse à la supervision des enregistrements audio-visuels par caméra dans des meeting-rooms. De tels studios sont appelés, parfois, des smart-

rooms (en Français : salle (*ou studio*) intelligente).

Ce travail de recherche se voit destiné à la supervision des multi-conférences et séminaires, utilisant l'information audio pour la localisation et le suivi du locuteur parlant.

En effet, nous tentons de réaliser et commander un système automatique pour orienter une ou plusieurs caméras mobiles ou fixes (*à focaliser*) vers des cibles sonores, représentées dans notre cas par des locuteurs fixes ou mobiles.

Les principaux objectifs de ce travail sont :

- d'une part, le développement et l'implémentation d'algorithmes d'identification, de localisation et de suivi des locuteurs actifs par des techniques d'ingénierie audio et la mise au point d'un système automatique de supervision, permettant le suivi du locuteur parlant (*locuteur actif*) par une caméra mobile ;
- et d'autre part, la participation à la formation de doctorants et thésards dans ce nouveau domaine de recherche.

Les étapes et taches du projet sont :

- 1- Mise à niveau et Formation ;
- 2- Acquisition de matériel spécifique ;
- 3- Enregistrement des bases de données audio-visuelles ;
- 4- Implantation des méthodes de détection et localisation sonore ;
- 5- Implantation des méthodes de contrôle de caméra ;
- 6- Réalisation de cartes électroniques et Expérimentation ;
- 7- Publications et Validation de notre système (et des résultats).

## INTRODUCTION

### Problématique

Les systèmes de télésurveillance, de vidéoconférence et de supervision des enregistrements audiovisuels utilisent des caméras fixes ou bien des caméras à orientation manuelle dirigées vers des sources sonores privilégiées [Lathoud, 2006] tels que, par exemple : un bruit de cassure, un conférencier actif ou un locuteur qui prend la parole.

Cette non-automatisation des systèmes existants a fait résulter un certain nombre de problèmes tels que : le surcoût engendré par les moyens matériels et humains employés et la fatigabilité des techniciens de supervision, puisque cette dernière exige qu'un caméraman (*au moins*) et des techniciens soit présent en permanence pendant les enregistrements.

### Solutions proposées

Pour remédier à ces inconvénients d'enregistrements audiovisuels manuels, nous pouvons proposer un système complètement automatique, basé sur une seule caméra mobile à orientation automatique vers la cible sonore [Khennouf, 2010] ou un système multi-caméras fixes avec switching automatique [Ouamour, 2002].

Notre application est destinée à la supervision automatique des enregistrements audiovisuels dans les meeting-rooms. Plusieurs locuteurs assis autour d'une table, prennent la parole à tour de rôle et le système de contrôle des caméras suit le locuteur parlant en toute harmonie.

Au préalable, différents types et différentes techniques de suivi des locuteurs ont été étudiés et expérimentés en off-line.

Pour ce faire nous avons proposé deux méthodes de localisation (*basées sur l'ingénierie audio*) : La première est basée sur la nouvelle technique de corrélation filtrée et la deuxième sur le différentiel d'énergie.

Pour la commande, nous avons proposé deux méthodes: l'une des méthodes est basée sur la commande par moteur pas à pas via un système automatique à réaliser. Une autre méthode est basée sur une nouvelle technique de commande virtuelle de caméra par le biais du contrôle de la ROI (*Region Of Interest*) [Khennouf, 2010].

Par ailleurs, certaines cartes électroniques d'interfaçage et de commande ont été réalisées.

### Objectifs

Notre application est destinée à la supervision automatique des enregistrements audiovisuels dans les meeting-rooms. Plusieurs locuteurs assis autour d'une table, prennent la parole à tour de rôle (*comme on a l'habitude de voir à la télévision*) et un système de caméra mobile ou fixe [Marchand, 2008] doit suivre le locuteur parlant.

Les objectifs sont multiples :

- Automatisation des studios d'enregistrement ;
- Diminution du surcoût du tournage (*prises de vues audiovisuelles par caméra*) ;
- Facilités apportées au metteur en scènes, lui permettant de diminuer l'effort physique et humain lors des prises de scènes ;
- Innovation technologique dans le domaine Multimédia, audio et audiovisuel ;
- Intérêt didactique et aide à la formation des étudiants de l'audio-visuel du partenaire économique.

## Description

Ce travail tient de l'automatisme, du traitement du signal audio et avec une certaine maîtrise des systèmes électroniques d'entrée-sortie.

Dans le cas d'une seule caméra mobile, on a proposé une tourelle mobile qui est composé d'un kit d'orientation automatique à base de rotor commandable électroniquement.

Pour le système mobile, une orientation dans le plan a été prévue : donc un degré de liberté. Nous nous intéressons au suivi des locuteurs positionnés autour d'une table ronde (*débats, meeting...*) par une caméra pivotante à réglage numérique et possédant une interface intelligente avec l'ordinateur.

Pour le système de commande, nous utiliserons un PC récent et rapide, pour permettre des temps de réponse courts et pour pouvoir se connecter à la caméra numérique (*port fire-wire*).

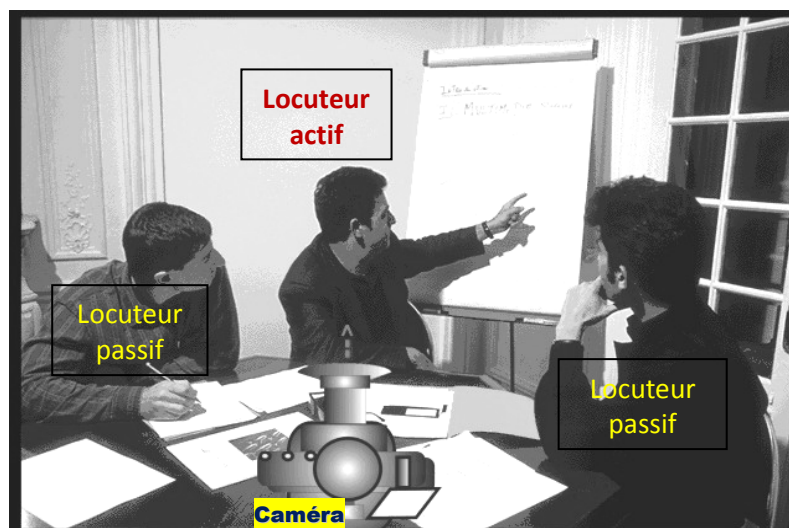


Fig.1 : Exemple de disposition circulaire :  
Suivi des interlocuteurs actifs positionnés autour d'une table ronde (*débats, meeting...*), par une caméra à localisation sonore.

Concernant les techniques de localisation des locuteurs, il s'agit d'une partie très consistante du travail de recherche. Pour les capteurs, on utilise des microphones. Ainsi, trois techniques de localisation sont déjà possibles :

- Technique à un seul microphone : utilisant les caractéristiques du signal sonore et l'approche d'identification [Ouamour, 2009];
- Technique invasive à deux microphones : utilisant une approche spatiale différentielle ;
- Technique non-invasive à deux microphones : utilisant l'approche du retard du signal audio.

Le suivant rapport est organisé comme suit : des généralités sur la localisation de source sonore sont données au chapitre I, La base de données réalisée est décrite au chapitre III, Les méthodes de localisation sonore proposées sont données au chapitre IV, le chapitre V énumère les systèmes proposés pour la commande de caméra, les résultats sont exposés au chapitre VI et une conclusion est donnée à la fin.

## CHAPITRE I

# GENERALITES SUR LA LOCALISATION DE SOURCE SONORE

### Introduction

Plusieurs méthodes de localisation sonore en Intelligence Artificielle s'inspirent du fonctionnement du système auditif humain pour la localisation d'une source sonore. L'objectif principal de ces méthodes, dites " bio-inspirées " est de réduire le système de perception à deux microphones seulement. Avant d'entamer une étude sur les méthodes de localisation sonore existantes, on débutera ce chapitre par une présentation des indices acoustiques utilisés par notre cerveau pour localiser un son.

### I.1 Aspects physiologiques du système auditif humain

Le rôle du système auditif de l'homme est de transformer un son en une suite d'activités nerveuses qui sont transmises sous formes de potentiels au cerveau. Ce dernier se charge de les traiter pour guider notre comportement. L'oreille, qui joue le rôle d'un transducteur mécanique en convertissant l'information de pression en une information électrique, est constituée de trois parties ; l'oreille externe, l'oreille moyenne et l'oreille interne [Purves, 2004] (*voir figure 1.1*).

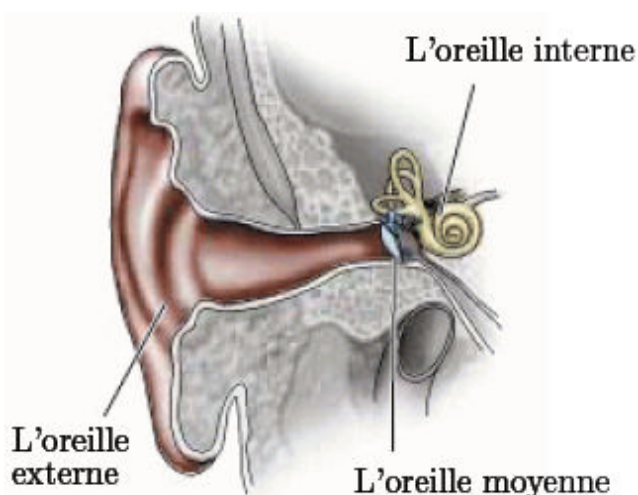


Fig. 1.1 : Système auditif humain

Avec une gamme de fréquences audibles s'étalant de 20 Hz à 20 kHz, l'oreille est capable de distinguer une variation de pression infime causée par un son de 1W avec une fréquence de 3

kHz à une distance d'environ 300km, dans un monde idéal complètement silencieux [Purves, 2004].

## I.2 Indices de localisation du son chez l'homme

L'homme est capable de localiser une source sonore avec une précision étonnante. Bien que cette précision varie selon les individus, la localisation des sons en provenance de l'avant est plus précise. L'erreur augmente jusqu'à  $20^\circ$  dans les autres directions. Les caractéristiques d'une onde acoustique sont modifiées par la présence de la tête avant leur interaction avec le pavillon de l'oreille. Ces modifications fournissent des indices acoustiques qui sont utilisées par le cerveau pour localiser une source sonore dans le plan azimutal et en élévation.

### I.2.1 Différence Interaurale en Amplitude

La découverte du premier indice acoustique utilisé pour la localisation horizontale est attribuée à Lord Rayleigh. Rayleigh a montré qu'un signal acoustique émis par une source sonore à partir de la droite d'un auditeur et perçu par son oreille droite, est plus intense que celui de l'oreille gauche. La différence des intensités de ces signaux (en dB) est appelée : Différence Interaurale en Amplitude, en anglais : Interaural Level Difference "ILD" (*voir figure 1.2*) [Hartmann, 1999].

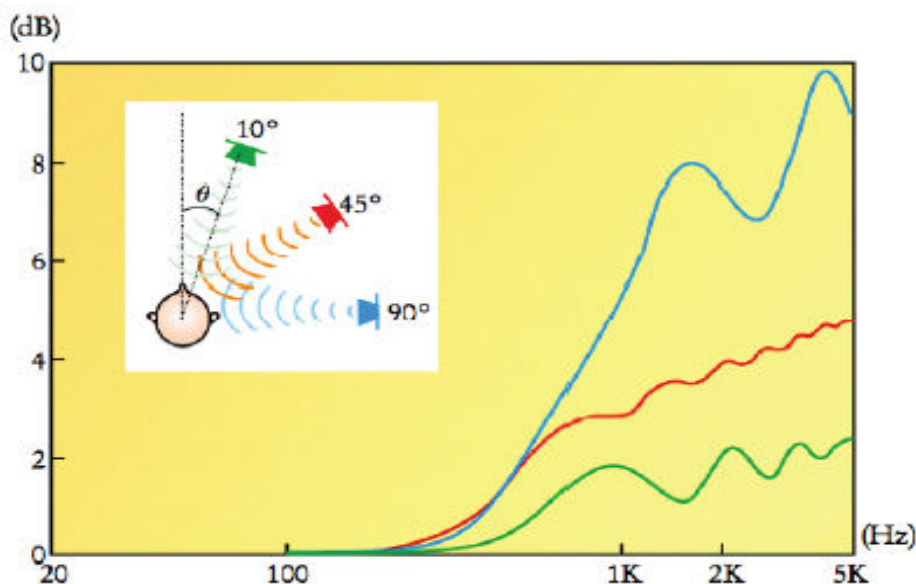


Fig. 1.2 : Différence Interaurale en Amplitude

### I.2.2 Différence Interaurale en Phase

Rayleigh a constaté aussi que l'homme est capable de détecter une différence en phase entre deux signaux perçus par nos oreilles. Ce deuxième indice, utilisé lui aussi dans la localisation



horizontale, est appelé ; Différence Interaurale en Phase, en anglais : Interaural Phase Difference " IPD ". Dans le cas d'un signal pluri-chromatique, cet indice est appelé ; Différence Interaurale Temporelle, en anglais : Interaural Time Difference " ITD ".

### I.2.3 Fonction de Transfert de la Tête

L'indice acoustique utilisé dans la localisation verticale est un indice spectral. Les ondes sonores sont modifiées considérablement par la présence des obstacles, y compris notre corps. Ces obstacles jouent le rôle de réflecteurs, modifiant ainsi les caractéristiques fréquentielles de ces ondes. Cet effet est assimilable à une opération de filtrage. Il est modélisé par la fonction de transfert anatomique appelé aussi fonction de transfert de la tête, en anglais : Head-Related Transfer Function " HRTF " (voir figure 1.3) [Hartmann, 1999].

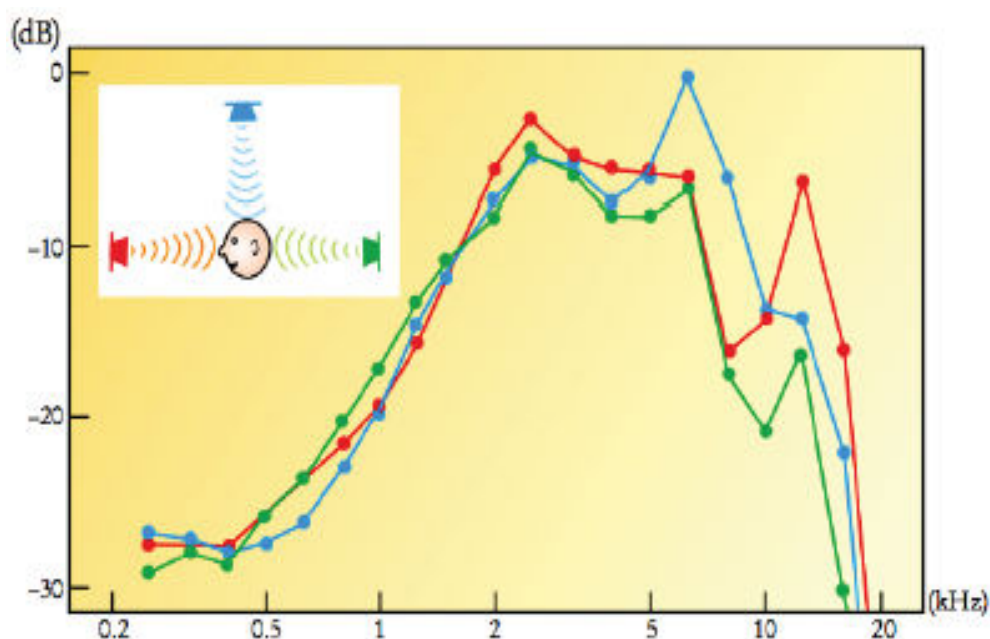


Fig. 1.3 : Fonction de Transfert de la Tête " HRTF "

### I.2.4 Perception de la distance et influence de l'environnement

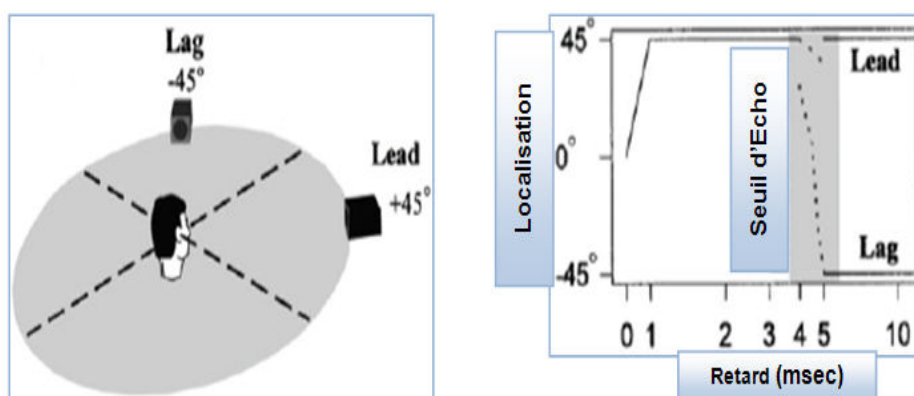
L'ensemble des indices acoustiques présentés précédemment, permet de déterminer très précisément l'azimut et l'élévation d'un son, mais en aucun cas la distance entre nos oreilles et la source sonore. La perception de cette distance n'est pas très bonne car elle est basée sur le niveau sonore émis par la source [Strybel, 1984]. La différence d'amplitude entre l'onde sonore reçue directement et celle ayant subi des réflexions peut constituer un bon indicateur

sur la distance à la source. Cependant, l'ensemble de ces réflexions vient perturber les autres indices acoustiques : l'IPD/ITD et ILD. Cette remarque est particulièrement vérifiée pour l'ITD, très sensible aux conditions environnementales puisqu'il est basé sur la cohérence des signaux perçus par les deux oreilles. Des essais acoustiques ont montré qu'il était probable, dans des environnements fortement réverbérant, que notre localisation soit basée uniquement sur l'ILD, plus robuste à de telles situations. Evidemment, cette stratégie est automatique et témoigne de nos capacités importantes d'adaptation à notre environnement sonore.

### I.2.5 L'effet de précedence

Un son atteint nos oreilles en parcourant plusieurs chemins. La première onde arrivée à nos oreilles joue un rôle fondamental dans notre perception. Ceci est expliqué par l'effet de précedence ou effet de Haas. Afin de mettre cet effet en évidence, des tests expérimentaux ont été effectués en chambre sourde avec deux haut-parleurs (Lead et Lag), situés à  $\pm 45^\circ$  et émettant un son court. Le signal émis par le premier haut-parleur "Lead" joue le rôle du chemin direct et le signal émis par le second haut-parleur "Lag" est retardé d'un délai variable, simulant ainsi l'écho.

Ces tests ont montrés qu'un écho parvenant à nos oreilles dans un délai compris entre 1ms et 5ms n'était pas perçu. Pour des délais supérieurs, les deux sons sont distingués comme deux sources différentes. Pour les délais inférieurs à 1ms, les deux sources sont fusionnées en une seule source située entre les deux haut-parleurs (*voir figure 1.4*) [Litovsky, 1999].



a) Disposition des deux haut-parleurs

b) Localisation en fonction du décalage

Fig. 1.4 : Illustration de l'effet de précedence

### I.3 Méthodes de localisation sonore en IA (*Intelligence Artificielle*)

Les méthodes utilisées en localisation sonore peuvent être classées selon deux grandes

catégories : les méthodes d'antennerie et les méthodes biomimétiques. Si l'utilisation des méthodes d'antennerie en robotique est particulièrement d'actualité, les méthodes biomimétiques restent l'intérêt d'une part importante de chercheurs.

### I.3.1 Les méthodes d'antennerie

Les méthodes d'antennerie sont basées sur l'utilisation des antennes de microphones. Depuis longtemps les antennes sont utilisées dans des applications diverses, mais leur exploitation en IA est restée assez limitée. Cependant, ces dernières années ont connues l'adoption progressive par la communauté robotique de méthodes éprouvées pour la localisation de source sonore.

#### I.3.1.1 La méthode haute résolution : MUSIC

La méthode MUSIC (MUltiple Signal Classification) est une méthode d'analyse spectrale paramétrique qui a permis une amélioration significative de la localisation d'une source sonore. Elle fait partie des méthodes dites "à haute résolution", dont elle est certainement la plus connue [Schmidt, 1979].

##### A. Principe de la méthode

On admet que la scène sonore comprend  $S$  sources à faible bande, de même fréquence, ainsi qu'une antenne de  $N$  microphones, avec  $S < N$ . Les sources sont supposées assimilables à des processus centrés, stationnaires et ergodiques (*voir figure 1.5*).

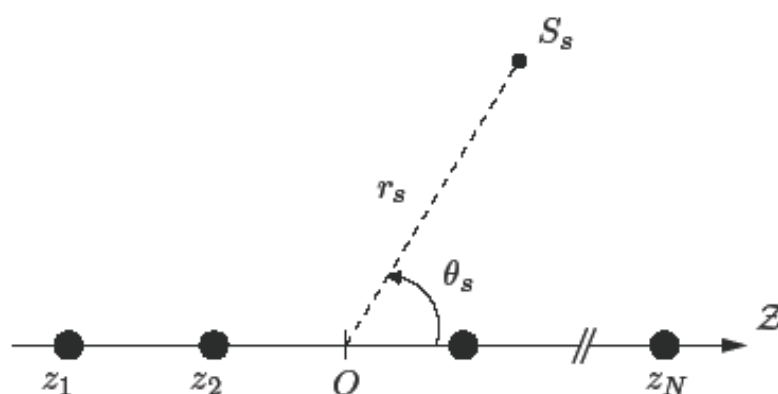


Fig. 1.5 : Organisation de l'antenne linéaire de microphones.

L'idée de la méthode MUSIC est donc d'exploiter le fait que les vecteurs d'antenne, aux directions des différentes sources prises séparément, sont orthogonaux à l'espace bruit déterminé par la décomposition spectrale de la matrice de covariance des signaux.

L'algorithme MUSIC est basé sur la décomposition en éléments propres de la matrice de covariance des signaux issus des microphones, aussi appelée matrice interspectrale. Il peut être synthétisé par les étapes suivantes :

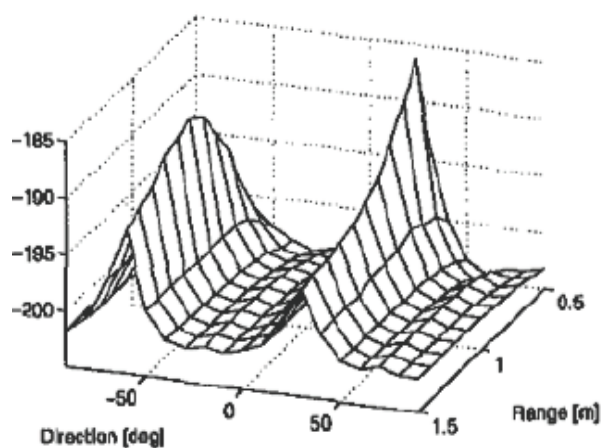
- *Acquisition des signaux  $m(t_1), \dots, m(t_L)$  par les microphones ;*
- *Calcul de l'estimée de la matrice de covariance (interspectrale) des signaux issus des microphones ;*
- *Décomposition de cette matrice en valeurs propres et vecteurs propres ;*
- *Séparation des espaces signal et bruit, et calcul du projecteur ;*
- *Définition du pseudo-spectre  $P(r)$ , sur un ensemble fini de valeurs de  $r$  ;*
- *Détection des positions en tant que maxima de  $P(r)$ .*

## B. Exploitation en IA

L'implémentation de cette méthode a été proposée dans [Asano, 1999], sur le robot Jijo-2, équipé de 8 microphones repartis sur sa circonférence. La localisation de la voix est effectuée en adaptant l'algorithme précédent au cas large bande. Les signaux reçus sur les microphones, échantillonnés à 16 kHz, sont décomposés dans le domaine fréquentiel par une FFT calculée sur 1024 points, à partir d'une fenêtre d'observation de 512 points, qui est décalée successivement de 160 points. L'estimation de la covariance s'effectue sur  $L=16$  valeurs, et conduit au résultat présenté sur la figure 1.6 suivante.



(a) Robot Jijo-2



(b) Résultats de localisation avec MUSIC

Fig. 1.6 : Application de la méthode MUSIC en robotique

### I.3.1.2 La méthode de corrélation

L'échantillonnage spatial opéré par une antenne de microphones est à l'origine de décalages

temporels entre les signaux disponibles en sortie des microphones appelés : " Time Delay Of Arrival " ou " TDOA ". Les méthodes de localisation sonore par corrélation visent à exploiter l'information spatiale incluse dans ces délais pour remonter à l'origine de la source sonore.

### A. Principe de la méthode

Pour ce faire, deux étapes sont indispensables :

- La première consiste à déterminer les TDOAs entre chaque paire de microphones ;
- La deuxième repose sur l'expression des TDOAs en fonction de la position de la source.

La méthode classique pour déterminer le décalage temporel entre deux signaux  $m_i(t)$  et  $m_j(t)$  perçus par deux microphones  $i$  et  $j$ , repose sur l'utilisation de la fonction d'intercorrélation. Les TDOAs correspondent aux valeurs des temps qui maximisent cette fonction [Argentieri, 2007].

Comme, la fonction d'intercorrélation dépend fortement des caractéristiques temporelles et fréquentielles des signaux, en particulier pour des signaux basses fréquences, les échantillons adjacents sont fortement corrélés, entraînant ainsi des "pics " pouvant être larges. Il s'avère donc que le choix de la durée d'observation  $T$  est critique quant aux performances de ces méthodes [Omologo, 1994].

[Silverman, 1992] suggère l'introduction de l'intercorrélation normalisée, en anglais : " Normalized Cross-Correlation " ou " NCC ". Cette normalisation par les énergies des signaux sur la fenêtre d'observation permet de faire ressortir sensiblement le pic de corrélation. Or, si la source sonore est périodique, des micropériodicités apparaissent dans la fonction d'intercorrélation normalisée.

[Omologo, 1994] rappelle les méthodes d'intercorrélation généralisée, en anglais : "Generalized Cross-Correlation" ou "GCC " qui adjoignent un terme de pondération fréquentielle discret, conduisant à une nouvelle fonction d'intercorrélation. Toute la difficulté réside donc dans le choix de la fonction de pondération.

[Knapp, 1976] propose différentes alternatives qui aboutissent à de multiples améliorations de la résolution de la fonction d'intercorrélation. La plus célèbre, appelée : Phase Transform ou " PHAT ", consiste à blanchir les spectres des signaux reçus au moyen de la fonction

définie par [Omologo, 1997].

[Valin, 2006] définit une nouvelle fonction de pondération pénalisant les fréquences où le rapport signal sur bruit est le plus faible, appelée : Reliability-Weighted Phase Transform (RWPCHAT).

Enfin que d'autres méthodes cherchent à améliorer les propriétés de la fonction d'intercorrélation en présence de réverbération. [Ma, 2004] propose une pondération adaptative conduisant à une meilleure robustesse vis-à-vis de ces perturbations que PHAT pour des temps de réverbération importants en environnement très bruité.

## **B. Exploitation en IA**

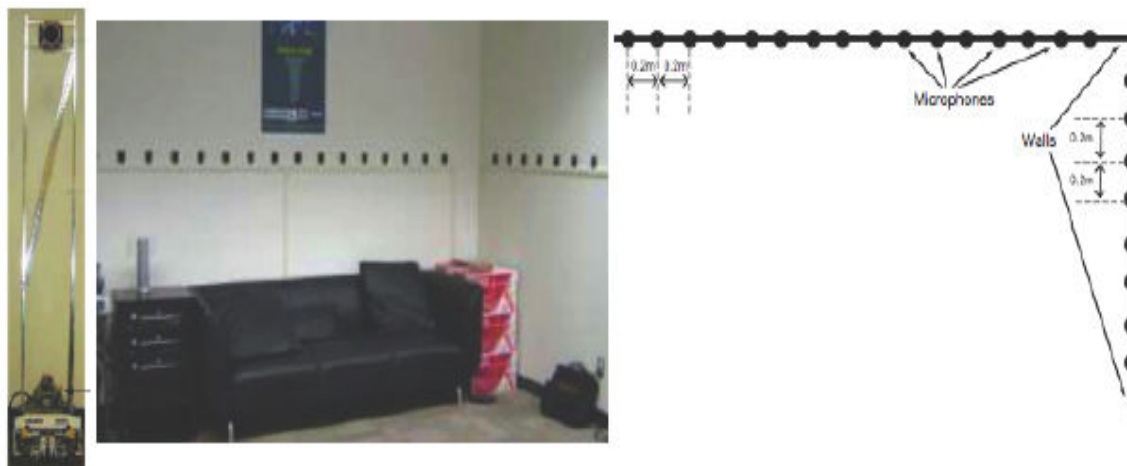
L'utilisation des méthodes avancées de corrélation est assez nouvelle en IA. Jusqu'à très récemment, le décalage temporel entre deux signaux a souvent été déterminé soit par la mesure des points de passage à zéro, soit par l'utilisation de la fonction d'intercorrélation classique.

[Cavaco, 1999] propose la décomposition fréquentielle des signaux au moyen d'un banc de filtres, puis de la détermination de l'ITD entre deux microphones par la mesure des points de passage à zéro.

[Okuyama, 2002] suggère l'exploitation de l'intercorrélation de façon à mesurer les TDOAs entre quatre microphones, positionnés sur les sommets d'un tétraèdre. L'originalité réside ici dans la sélection de la fenêtre d'observation : plutôt que de calculer l'intercorrélation sur la totalité des signaux, un seuil permet de détecter les zones temporelles libres d'écho pendant lesquelles les TDOAs sont déterminés.

Dans cette même approche, [Murray, 2005] calcule le TDOA entre deux microphones à partir d'une simple corrélation pour déterminer l'azimut d'une source sonore.

[Wang, 2004] utilise le PHAT sur une antenne de microphones disposés sur les murs d'une pièce devant être visitée par un robot-guide. Le but est de localiser ce robot grâce au discours qu'il émet depuis un haut-parleur. La localisation sonore est alors fusionnée avec les informations perçues par ses capteurs proprioceptifs, de façon à le recalculer sur la carte de l'environnement (*voir figure 1.7 suivante*).



(a) Robot-guide      (b) pièce à visiter      (c) Détails sur l'antenne de microphone

Fig. 1.7 : Présentation du robot-guide et disposition de l'antenne de microphone.

### I.3.1.3 La méthode de formation de voie

Les approches par formation de voie sont probablement les plus utilisées en robotique. Leur simplicité et leur faible coût calculatoire les rend particulièrement bien adaptées à ce contexte. Cependant, leurs performances en localisation dépendent très fortement de la taille et du nombre de microphones utilisés.

#### A. Principe de la méthode

La formation de voie est une méthode qui permet la focalisation d'une antenne de microphones dans une direction spécifique de l'espace. Les signaux sonores provenant d'une direction spatiale d'intérêt sont ainsi amplifiés, tout en atténuant les autres directions d'incidence, afin de localiser ou d'isoler (extraire) une ou plusieurs sources sonores.

Cette méthode exploite les signaux  $m_1(t), \dots, m_n(t)$  des  $N$  microphones d'une antenne en les combinant linéairement à l'aide de filtres de réponse impulsionnelle  $w_n(t)$ , pour obtenir un signal de sortie  $y(t)$  résultat d'une sommation constructive des signaux (voir figure 1.8).

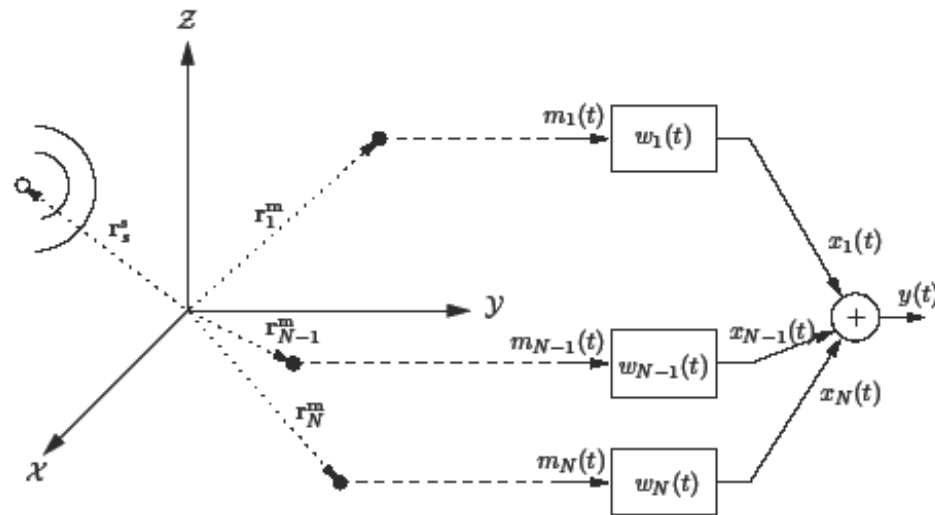


Fig. 1.8 : Principe de la formation de voie

Nous pouvons donc écrire :

$$y(t) = \sum_{n=1}^N w_n(t) * m_n(t) \quad (1.1)$$

avec  $*$  est un produit de convolution.

Une formation de voie est caractérisée par sa réponse spatio-fréquentielle. Cette fonction des variables d'espace et de la fréquence peut être assimilée à une fonction de transfert normalisée entre une source située en  $\mathbf{r}$  et la sortie  $y(t)$ , et indique quelles zones de l'espace sont amplifiées ou atténuées. Plusieurs observations utiles pour une exploitation en robotique peuvent être établies en traçant le diagramme d'antenne dans les configurations suivantes :

- Variation du nombre de microphones  $N$ , la longueur de l'antenne  $L$  et la fréquence  $f$  étant fixées ;
- Variation de la longueur  $L$ , pour  $N$  et  $f$  constants ;
- Variation de la fréquence  $f$  pour  $N$  et  $L$  constants.

Une telle étude, très classique, est menée dans [McCowan, 2001].

## B. Exploitation en IA

Les propriétés du filtrage spatial peuvent être utilisées afin de localiser une source sonore dans l'environnement. Le principe commun en IA est extrêmement simple : il consiste à polariser l'antenne de microphones dans chacune des directions à explorer, afin d'obtenir une carte de puissance acoustique, qui présente un pic dans la direction d'émission de la source



sonore à localiser. La construction de cette carte nécessite de former le diagramme d'antenne pour qu'il soit sensible à une direction.

Ainsi, bien que [Tamai 2004] a utilisé une antenne, équipée de 128 microphones répartis dans une pièce, la mauvaise focalisation pour les grandes longueurs d'onde l'a obligé à filtrer les fréquences inférieures à 500 Hz (voir figure 1.9).

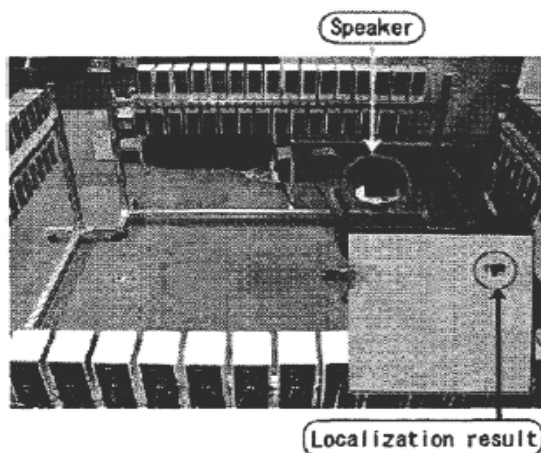


Fig. 1.9 : Localisation sonore à partir d'une antenne de 128 microphones

[Mattos, 2004] arrive à la même conclusion par simulation concernant l'antenne implémentée sur la petite plateforme mobile présentée dans la figure 1.10.



Fig. 1.10 : Robot EvBot-II équipé de son antenne à 8 microphones

Les 8 microphones sont insérés sur une ceinture entourant le robot, mais leur emplacement est décidé de manière empirique à partir d'un logiciel de simulation développé pour tracer la réponse d'antenne. Celle-ci laisse paraître un lobe principal fin pour les fréquences supérieures à 1kHz, tandis que les fréquences inférieures à 800Hz demeurent inexploitable pour la localisation.

## CHAPITRE II

### BASE DE DONNEES REALISEE

#### **Introduction**

Pour les tests d'évaluation de notre système, nous avons conçu plusieurs bases de données qui sont constituées de plusieurs scénarios avec différentes configurations. Les enregistrements audio ont été faits à l'aide d'un système d'acquisition spécialisé.

#### **II.1 Scénarios des bases de données**

Les scénarios qui constituent les bases de données peuvent être classés dans trois catégories : La première catégorie contient un seul locuteur qui peut être masculin ou féminin. Dans la deuxième, on trouve des dialogues entre deux locuteurs masculins ou entre un locuteur masculin et un locuteur féminin. La troisième catégorie contient un discours entre plusieurs interlocuteurs masculins et féminins (trois locuteurs) qui prennent la parole alternativement dans une manière naturelle.

#### **II.2 Configuration des bases de données**

Nous avons proposé deux configurations principales pour les différents scénarios : Une configuration fixe et une configuration mobile. Dans la configuration fixe le locuteur est assis devant une table dans une des positions fixes : gauche, milieu ou droite, par exemple. Cependant, dans la configuration mobile, le locuteur se déplace lentement d'une extrémité vers une autre.

#### **II.3 Discours enregistrés**

On a utilisé une variété de discours pendant les enregistrements des bases de données tels que : des sons phonémiques, des paragraphes, des discours et des conversations de manière naturelle.

## II.4 Conception de la base de données

Ainsi, on a effectué plusieurs manipulations pour construire une base de données riche et équilibrée. Nous avons changé plusieurs paramètres dans un but de réaliser une base de données significative avec de bonnes performances. Ces paramètres sont illustrés dans la figure 2.1 ci-dessous.

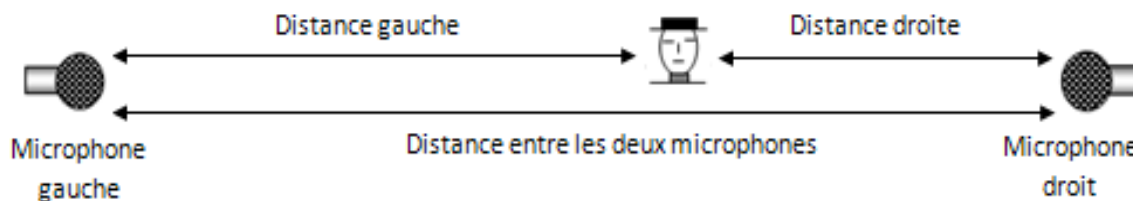


Fig. 2.1 : Les différents paramètres des expériences réalisées

La description de la base de données conçue est donnée ci-après, en Anglais. Celle-ci peut être retrouvée aussi sur notre site web : <http://sayoud.net/ASDB-2.pdf>

**ASDB-2 ASDB-2 ASDB-2 ASDB-2 ASDB-2**

**Version descriptive en Anglais**

### INTRODUCTION

This new stereophonic speech database, called 2-channels **Audio Signals DataBase (ASDB-2)**, was recorded in February 2013 at the **USTHB** University in the framework of our PNR research project.

### PURPOSE OF USE

This stereophonic speech corpus has been built for a purpose of speaker localization (in 1 dimension). It could be used for a task of speaker recognition. It could be used for a task of speech recognition too.

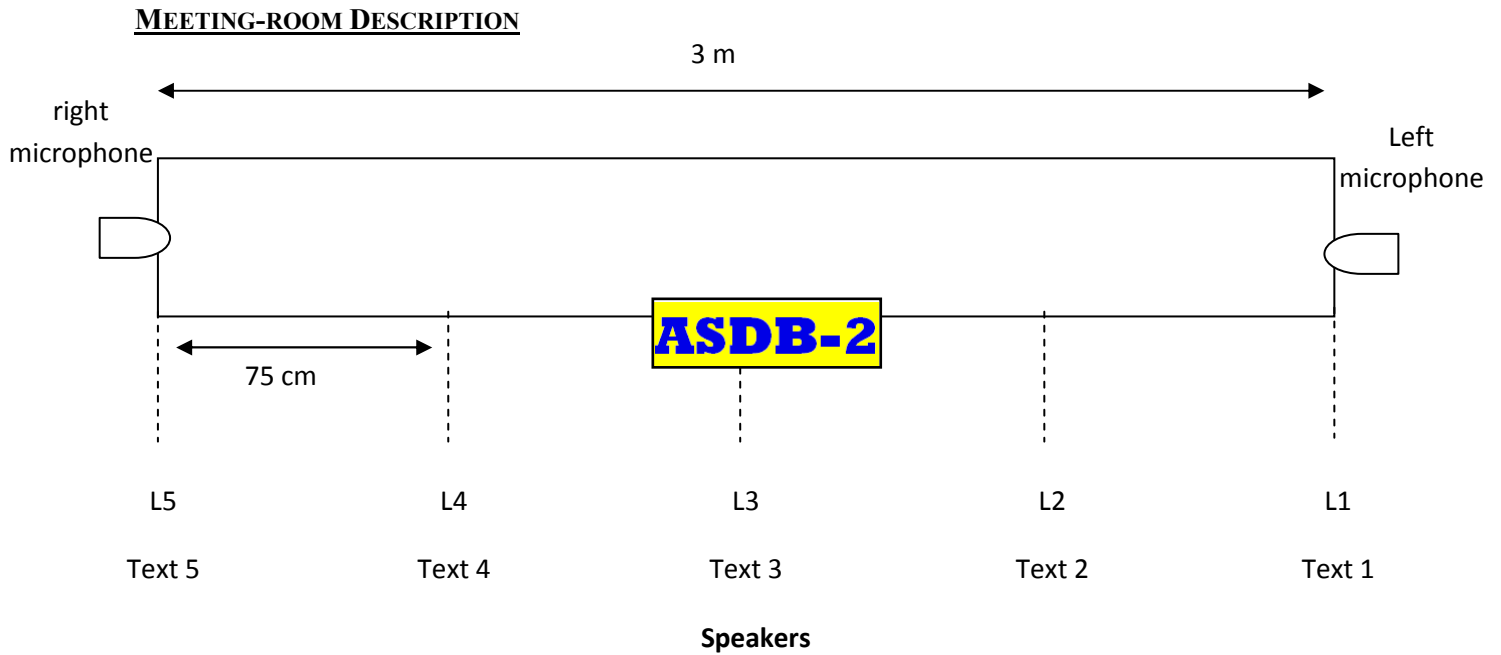


Fig. 1: Photo of the meeting-room (USTHB Bab-Ezzouar)



Fig. 2: Photo of the anechoic-room (USTHB Bab-Ezzouar)



Fig. 3: Photo of the right cardioid microphone: Schure beta-58s (USTHB Bab-Ezzouar)



Fig. 4: Photo of a configuration with 3 speakers (USTHB Bab-Ezzouar).

### **CORPUS DESCRIPTION**

There are 13 speakers: 5 female and 8 male speakers.

Language: French (most of the time) and Arabic (rarely)

The microphones are Shure 58s-Beta

The sampling rate is about 44kHz with a length of 16 bits per channel.

The 5 different texts (Text 1, Text 2, ... Text 5) are displayed below:

### **SCENARIOS DESCRIPTION**

There are 4 folders in the DVD:

- Fixed Positions folder: containing 5 scenarios per speaker (L1, L2 ... L5); each scenario corresponds to one position (L1, L2, .. L5) as displayed on the above figure. Totally, there are 13 speakers, then:  $13 \times 5 = 65$  different scenarios in this folder. In position L<sub>j</sub>, the speaker reads the text Text j.
- Multi Speakers folder: 3 different speakers are sitting at 3 different positions in the meeting-room. They speak sequentially in a form of a debate or discussion. L: stands for the extreme left, R: stands for the extreme right, M: stands for the middle position, RM stands for intermediate middle between R and M, and LM stands for intermediate middle between L and M.
- Moving Speakers folder: speakers are moving from one side to the other (from the left microphone to the right one) by speaking continuously, without stopping.

- Anechoic room folder: 5 different free scenarios are recorded in an anechoic room. There are 5 different speakers in 5 fixed positions of the meeting-room, who speak sequentially in a form of free discussion or debate.

## TEXT DESCRIPTION

### Text 1

*« Le traitement du signal est la discipline qui développe et étudie les techniques de traitement, d'analyse et d'interprétation des signaux. Parmi les types d'opérations possibles sur ces signaux, on peut dénoter le contrôle, le filtrage, la compression de données, la transmission de données, le débruitage, la déconvolution, la prédiction, l'identification, la classification, etc.*

*Bien que cette discipline trouve son origine dans les sciences de l'ingénieur (particulièrement l'électronique et l'automatique), elle fait aujourd'hui largement appel à de nombreux domaines des mathématiques, comme la théorie du signal, les processus stochastiques, les espaces vectoriels et l'algèbre linéaire et des mathématiques appliquées, notamment la théorie de l'information, l'optimisation ou encore l'analyse numérique. »*

### Text 2

*« Le traitement peut être fait, sans numériser les signaux, par des circuits électroniques analogiques ou aussi des systèmes optiques (traitement du signal optique).*

*Il est de plus en plus souvent réalisé par traitement numérique du signal, à l'aide d'ordinateurs, de microprocesseurs embarqués, de microprocesseurs spécialisés nommés DSP, de circuits reconfigurables (FPGA) ou de composants numériques dédiés (ASIC).*

*Il existe plusieurs branches particulières du traitement du signal, en fonction de la nature des signaux considérés. En particulier :*

- Traitement de la parole (ou plus généralement du son) - pour l'analyse, la compression, la synthèse et la reconnaissance de la parole*

- *Traitement d'images - pour l'analyse, la restauration et la compression d'images fixes* »

### Text 3

« *On utilise différentes mesures représentatives de la qualité d'un signal et de l'information contenue :*

*Le rapport signal sur bruit, notion utilisée très fréquemment mais équivoque puisque tout dépend de ce qui est considéré comme signal et comme bruit.*

*Le nombre de bits effectifs Effective Number of Bits (ENOB) qui est une mesure de la qualité de conversion analogique-numérique.*

*L'information de Fisher, utile en particulier en estimation de paramètres. Elle peut être relative à un seul ou plusieurs paramètres (matrice d'information de Fisher).*

*L'entropie, grandeur issue de la physique statistique et de la théorie de l'information (travaux de Shannon), utilisée dans les opérations de codage. »*

### Text 4

« *Le traitement du son s'est largement amélioré grâce aux ordinateurs. Toutefois, certains musiciens parlent davantage d'un son de nature différente [ ] que d'une simple amélioration qualitative (de même que le CD ne « sonne » pas comme le disque vinyle, et que certains groupes, par exemple Genesis, ont particulièrement profité du « nouveau son » offert par le nouveau support). La synthèse sonore permet en outre de créer des sons artificiels ou de recréer les sons d'instruments naturels. Elle a été à l'origine de nombreux bouleversements en musique.*



- l'analyse des échos permet d'obtenir des informations sur le milieu sur lequel les ondes se sont réfléchies. Cette technique est exploitée dans le domaine de l'imagerie radar ou sonar. »

### Text 5

« Le traitement du signal est une discipline qui est méconnue car elle se fonde à l'intersection de plusieurs domaines scientifiques. Les premiers développements trouvent leurs origines dans l'électronique et l'automatique. Dans les années 1960, l'arrivée des ordinateurs et le début de la numérisation des signaux font basculer la discipline dans l'informatique moderne.

Mais avant tout, le traitement du signal tire ces outils de plusieurs domaines des mathématiques, dont les résultats ont été formulés un siècle avant ! On peut ainsi évoquer les espaces vectoriels, l'algèbre linéaire, les probabilités, les statistiques, les distributions, l'optimisation,... En 1949, Shannon théorise le concept d'information très utile en traitement du signal. »

**THE RESEARCH PROJECT PNR TEAM : A. Houacine, S. Mekaoui, S. Khenouf, S. Ouamour and H. Sayoud.**

*USTHB University.*

*ASDB-2 2013*

<http://sayoud.net/ASDB-2.pdf>

## CHAPITRE III

### METHODES DE LOCALISATION SONORE PROPOSEES

#### Introduction

Nous avons proposé deux nouvelles méthodes pour la localisation des sources sonores fixes et mobiles. La première méthode est basée sur la corrélation que nous avons appelée : méthode de la corrélation filtrée, en anglais : Filtred Correlation Method ou FCM, tandis que la deuxième méthode est basée sur l'énergie des deux signaux perçus par les deux microphones gauche et droit, que nous avons appelée : Méthode du différentiel d'énergie, en anglais : Energy Differentiel Method ou EDM.

#### III.1 Corrélation temporelle

La fonction de corrélation est définie par la mesure de similitude entre deux signaux, dont l'un est décalé dans le temps par rapport à l'autre, en fonction du décalage entre les deux signaux. On distingue deux types de corrélation : l'autocorrélation et l'intercorrélation (Cross-correlation).

##### III.1.1 Autocorrélation temporelle

L'autocorrélation consiste à comparer un signal avec lui-même décalé d'une certaine valeur  $\tau$  durant un intervalle de temps.

La fonction d'autocorrélation temporelle est donnée par la valeur moyenne temporelle du produit de  $x(t)$  par  $x(t + \tau)$  [Kunt, 1996] :

$$Cor_{xx}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t + \tau) dt \quad (3.1)$$

##### III.1.2 Intercorrélation temporelle

L'intercorrélation consiste à comparer deux signaux différents dont l'un est décalé d'une certaine valeur  $\tau$ .

La fonction d'intercorrélation temporelle est donnée par la valeur moyenne temporelle de

produit de  $x(t)$  par  $y(t + \tau)$ .

$$Cor_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) y(t + \tau) dt \quad (3.2)$$

### III.2 Méthode de la Corrélation Filtrée (FCM)

#### III.2.1 Calcul de la corrélation Discrète

La corrélation discrète est calculée pour chaque paire de segments de parole de durée de 0,25 secondes pour les signaux perçus par les deux microphones : Le signal  $x$  perçu par le microphone droit et le signal  $y$  perçu par le microphone gauche. Les étapes de calcul de la corrélation filtrée sont décrites comme suit :

On calcule d'abord la corrélation discrète par :

$$Cor_{xy}(m) = \sum_{i=-N}^{+N} x(i) \cdot y(i + m) \quad (3.3)$$

Ensuite, on applique au signal de corrélation discrète ( $Cor_{xy}$ ) un filtre passe-bande elliptique, avec des fréquences de coupure de 20Hz et 300Hz. Le signal de corrélation filtrée obtenu est nommé  $FCor_{xy}$ .

#### III.2.2 Localisation de la position par la méthode FCM

En considérant un locuteur entre deux microphones distants (*voir figure 3.1*), et en supposant que les distances entre le locuteur et les deux microphones sont comme suit :

- $d_r$  : La distance entre le locuteur et le microphone droit ;
- $d_l$  : La distance entre le locuteur et le microphone gauche ;
- $L$  : La distance entre les deux microphones.

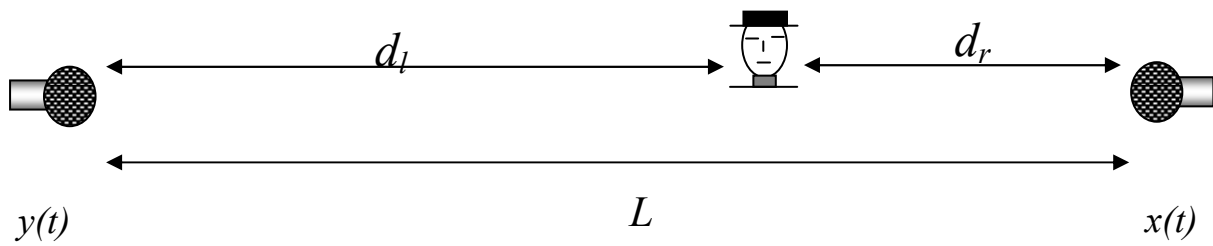


Fig. 3.1 : Disposition du locuteur entre les deux microphones

Alors on peut dire :

$$L = d_l + d_r \quad (3.4)$$

$$\Delta = d_r - d_l \quad (3.5)$$

$$T_{\max} \text{Cor} = \frac{d_r - d_l}{c} \quad (3.6)$$

On a donc:

$$d_r + d_l = c.T_{\max} \text{Cor} \quad (3.7)$$

Et

$$d_r = (c.T_{\max} \text{Cor} + L) / 2 \quad (3.8)$$

Le son se propage à une vitesse approximative de 340 m/s dans l'air à une température de 15°C, d'où:

$$d_r = (340.T_{\max} \text{Cor} + L) / 2 \quad (3.9)$$

### III.2.3 Exemple de calcul de la distance $d_r$

Si la fréquence d'échantillonnage est :  $f_e = 44 \text{ kHz}$  et  $m_{\max} = 30$

qui correspond à une valeur de :

$$T_{\max} \text{Cor} = 30 / 44 \text{ kHz} \quad (3.10)$$

et si  $L = 30 \text{ cm}$

donc, on a :  $d_r = 26,6\text{ cm}$

### III.2.4 Calcul des distances $d_r$ et $d_l$ en pratique

En pratique la distance du microphone droit  $d_r$  et la distance du microphone gauche  $d_l$  sont calculées comme suit :

$$d_{r\_practice} = d_r + \text{delta} \quad (3.11)$$

et

$$d_{l\_practice} = d_l + \text{delta} \quad (3.12)$$

où  $\text{delta}$  est un petit nombre correctif qui a été déduit expérimentalement :

$$\text{delta} = -0,2 \quad (3.13)$$

### III.2.5 Résultats techniques obtenus

Ci-dessous quelques résultats technique obtenus pendant les expériences de localisation de position réalisées avec la methode FCM:

#### Exemple 1 :

Dans cet exemple, nous avons calculé  $T_{\max}$  en fonction du temps de déplacement d'un locuteur qui se deplace lineairement avec une faible vitesse de la droite vers la gauche (voir figure 3.2).

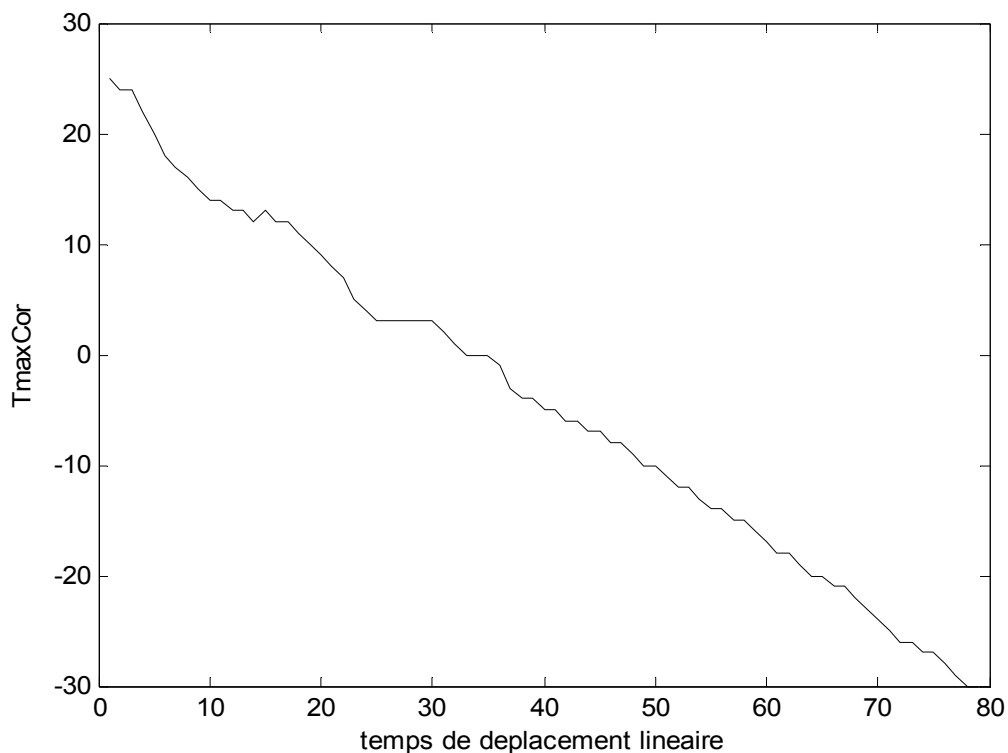


Fig. 3.2 :  $T_{\max}$  en fonction du temps de déplacement linéaire d'un locuteur qui se déplace d'une extrémité à une autre (de la droite vers la gauche).

### Exemple 2 :

Dans cet exemple, on procède à la localisation (position par rapport au microphone gauche en mètre) par la méthode FCM d'un locuteur se déplaçant linéairement avec une vitesse constante, partant du microphone gauche vers le microphone droit et en prononçant des phonèmes. La distance entre les deux microphones est 2m (voir figure 3.3).

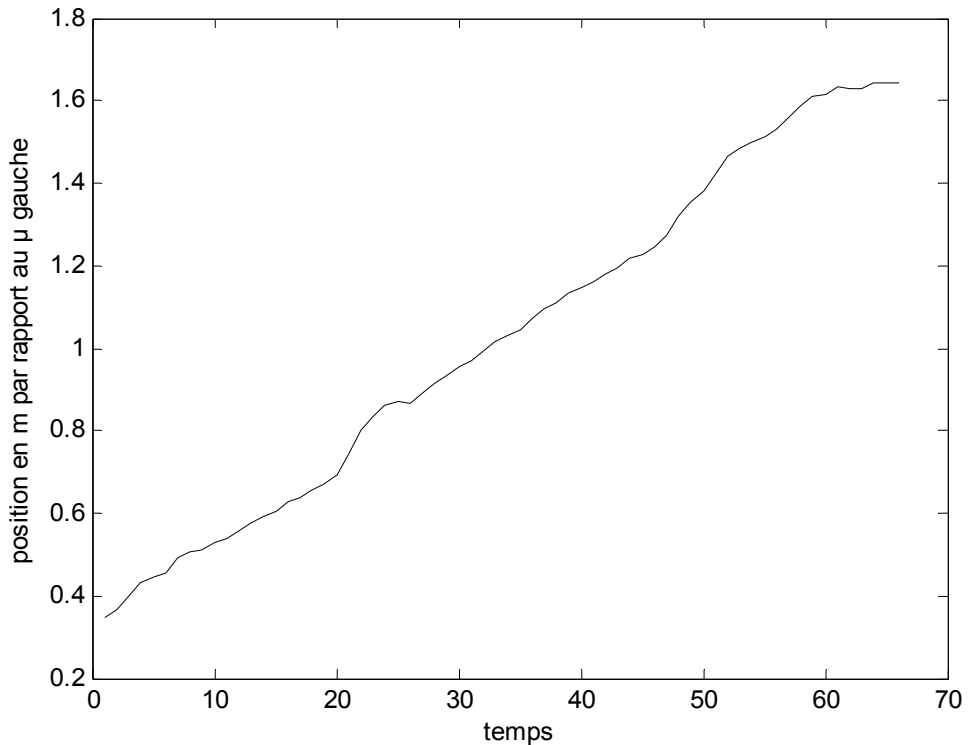


Fig. 3.3 : Position (par rapport au microphone gauche en m) d'un locuteur en déplacement linéaire avec une vitesse constante, partant du microphone gauche au microphone droit en prononçant un phonème.

### III.3 Méthode du différentiel d'énergie (EDM)

#### III.3.1 Calcul du différentiel d'énergie

Dans notre application, on a utilisé l'énergie modifiée de norme L1 (en db/20) qui est calculée comme étant la somme des valeurs absolues des valeurs du signal, le long de la fenêtre d'analyse. Cette énergie peut être appelée : Energie d'ordre 1.

Les deux énergies ; l'énergie du signal perçu par le microphone droit et l'énergie du signal perçu par le microphone gauche, sont données par :

$$E_x = \sum_{i=0}^N |x_i| \quad (3.14)$$

$$E_y = \sum_{i=0}^N |y_i| \quad (3.15)$$

Le différentiel de l'énergie logarithmique  $DE_{xy}$  est donné par :

$$DE_{xy} = \log(E_x/E_y) = \log(E_x) - \log(E_y) \quad (3.16)$$

De la même manière la somme des énergies logarithmiques  $SE_{xy}$  est donnée par :

$$SE_{xy} = \log(E_x \cdot E_y) = \log(E_x) + \log(E_y) \quad (3.17)$$

### III.3.2 Détecteur d'activité

Avant de procéder à la localisation d'un locuteur on doit chercher à détecter s'il y a une activité ou pas (cas de silence). Cette procédure est appelée détection d'activité et de silence, en anglais : Silence Activity Detection (SAD). Pour ce faire on compare notre signal à un seuil qu'on doit fixer au préalable à une valeur expérimentale [Sayoud, 2003<sup>T</sup>]. L'algorithme suivant assure cette procédure :

**Si**  $SE_{xy} < \text{Seuil}$

*donc silence*

**Sinon** *activité*

### III.3.3 Localisation de la position par la méthode EDM

Maintenant, il est facile d'estimer la position relative du locuteur par rapport à la position des microphones. Par exemple, il est facile de déduire si le locuteur est du côté gauche, du côté droit ou tout simplement au milieu, par l'algorithme suivant :

*Calcul du différentiel d'énergie;*

**Si**  $DE_{xy} < \text{Threshold}_{min}$

*donc locuteur est à gauche*

**Si**  $DE_{xy} > \text{Threshold}_{max}$

*donc locuteur est à droite*

**Si**  $> \text{Threshold}_{min}$  et  $DE_{xy} < \text{Threshold}_{max}$

*donc locuteur est au milieu*

*$\text{Threshold}_{min}$  et  $\text{Threshold}_{max}$  sont réglés expérimentalement.*

### III.3.4 Résultats techniques obtenus

Ci-dessous quelques résultats techniques obtenus pendant les expériences de localisation de



position par la méthode EDM :

### Exemple 1 :

Dans cet exemple, on procède à la localisation par la méthode EDM d'un locuteur se déplaçant linéairement avec une vitesse constante, et en partant du microphone gauche au microphone droit (voir figure 3.4).

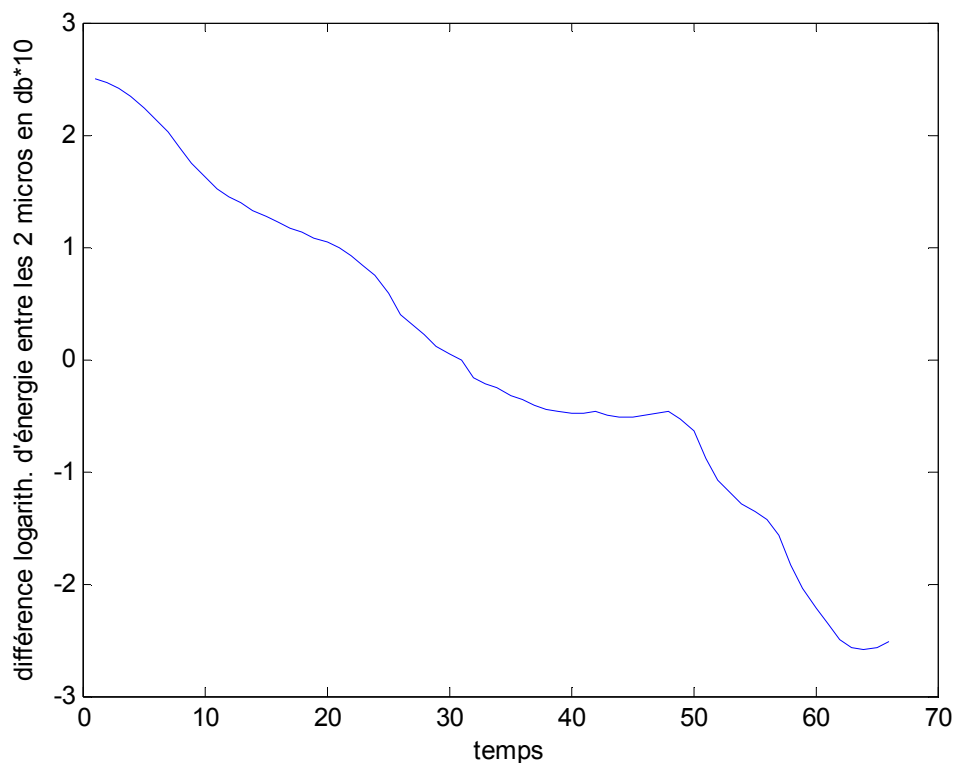


Fig. 3.4 : Energie différentielle d'un locuteur se déplaçant linéairement à vitesse constante du microphone gauche vers le microphone droit en prononçant un phonème.

### Exemple 2 :

Dans cet exemple, on procède à la localisation par la méthode EDM d'un locuteur se déplaçant pas à pas en s'arrêtant chaque 20cm, en partant du microphone gauche au microphone droit et en prononçant des phonèmes (voir figure 3.5).

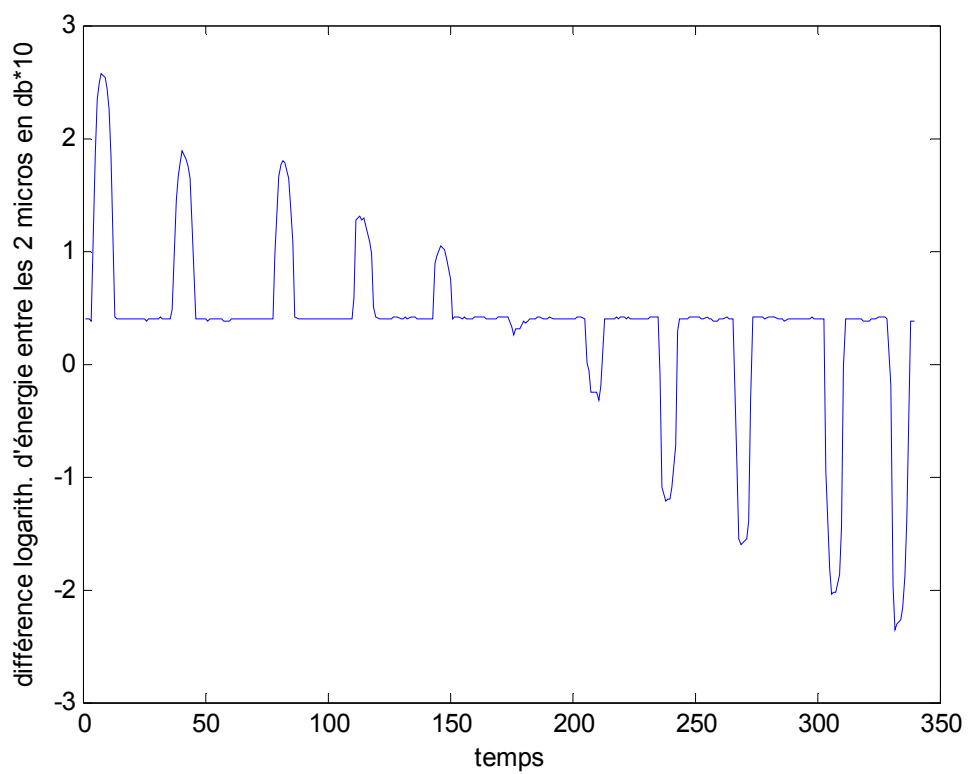


Fig. 3.5 : Différence logarithmique d'énergie d'un locuteur, qui se déplace pas à pas en s'arrêtant chaque 20cm, du microphone gauche vers le microphone droit et en prononçant des phonèmes.

## CHAPITRE IV

### SYSTEMES PROPOSES POUR LA COMMANDE DE CAMÉRA

#### Introduction

Une fois la localisation du locuteur actif est faite, en utilisant une des deux méthodes de localisation proposées (EDM ou FCM), on procède alors à la génération du signal de commande qui va être utilisé par le système de commande pour piloter l'orientation de la caméra vers le locuteur actif détecté. Nous avons conçu deux systèmes de commande : un système d'orientation mobile à base de moteur pas à pas et un nouveau système de commande virtuel.

#### IV.1 Schéma synoptique du système automatique de commande

L'objectif de ce travail est la conception d'un système automatique qui assure la tâche du suivi d'une cible sonore (un locuteur, dans notre application). Les locuteurs sont assis derrière une table comme illustré dans la figure 4.1 suivante :

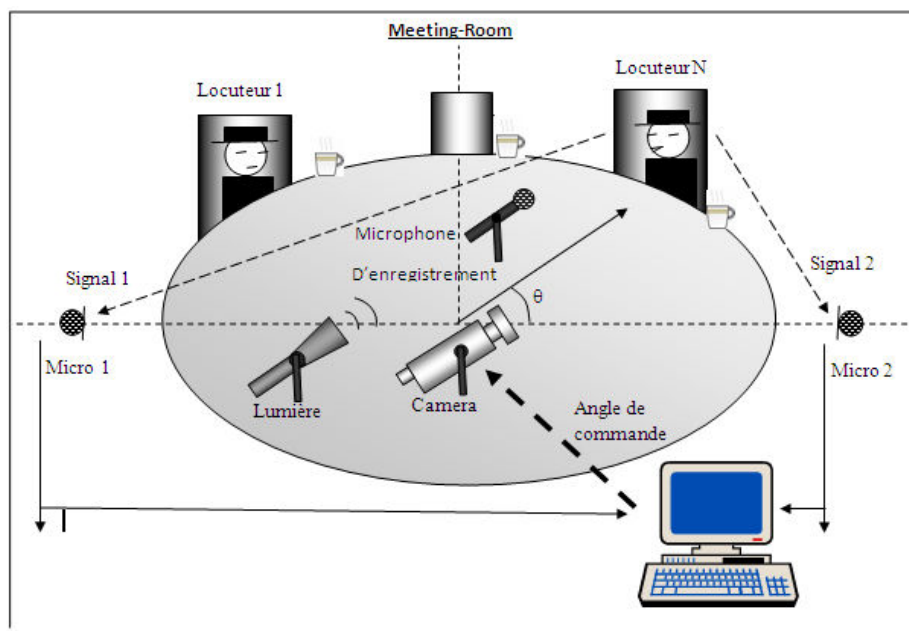


Fig. 4.1 : Schémas Synoptique du système automatique de commande.

#### IV.2 Système d'orientation mobile de la caméra

Le système d'orientation mobile est constitué d'un micro-ordinateur, d'un module de puissance et d'un support motorisé. Le support motorisé contient une base en acier, un moteur pas à pas et un axe en aluminium sur lequel est montée la caméra.

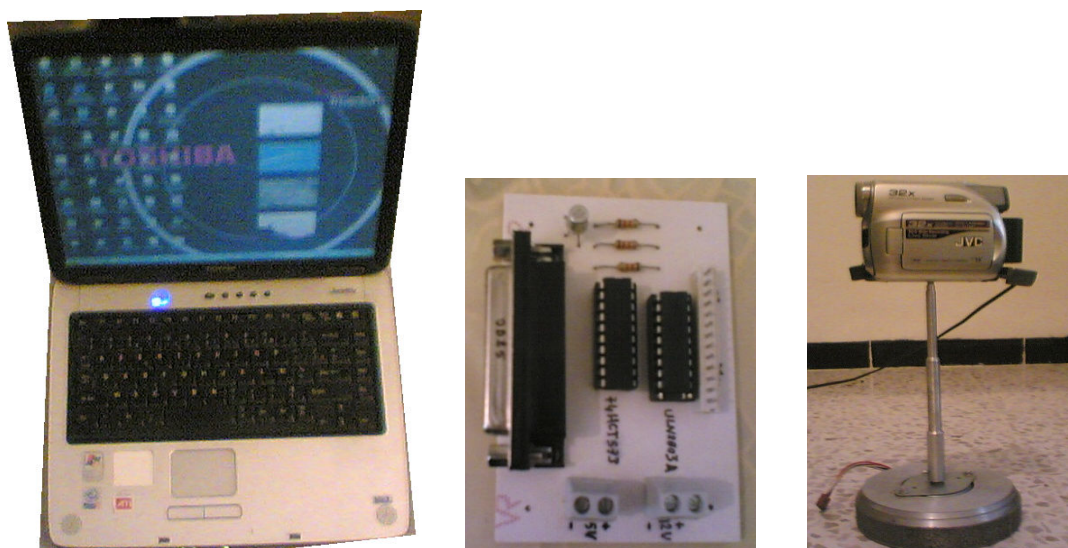


Fig. 4.2 : Les trois modules constituant le système d'orientation de la caméra.

## IV.2.1 Génération des signaux de commande

Les signaux de commande sont générés à partir des signaux de localisation de position en utilisant la méthode basée sur l'énergie (EDM) ou la méthode basée sur la corrélation (FCM). La commande est représentée par l'orientation de la caméra vers les différentes positions. Exemple, pour 3 positions, on a : la gauche (-1), la droite (+1) et le milieu (0).

### IV.2.1.1 Algorithme

La génération des signaux de commande est faite selon un algorithme qui est illustré ci-dessous :

Exemple d'algorithme pour 3 positions :

***Pour temps=début : Fin,***

***Si Silence donc :***

***Commande <sub>Caméra</sub> = Milieu;***

***Fin Si***

***Fin Pour***

***Pour temps=début : Fin,***

***Si Position détectée  $\neq$  Silence,***

***Si Position détectée  $\neq$  Position précédente,***

***Commande <sub>Caméra</sub> = Position détectée;***

***Sinon***

***Commande <sub>Caméra</sub> = Commande <sub>Caméra</sub> précédente;***

***Fin Si***

***Sinon***

***Commande <sub>Caméra</sub> = Commande <sub>Caméra</sub> précédente;***

***Fin Si***

***Fin Pour***

#### **IV.2.1.2 Exemples de résultats obtenus**

On appliquant l'algorithme précédent aux différents scénarios à position fixe, on obtient des signaux intermédiaires qui serviront à la fabrication du signal de commande final. En voici quelques uns dans la figure 4.3 ci-dessous.

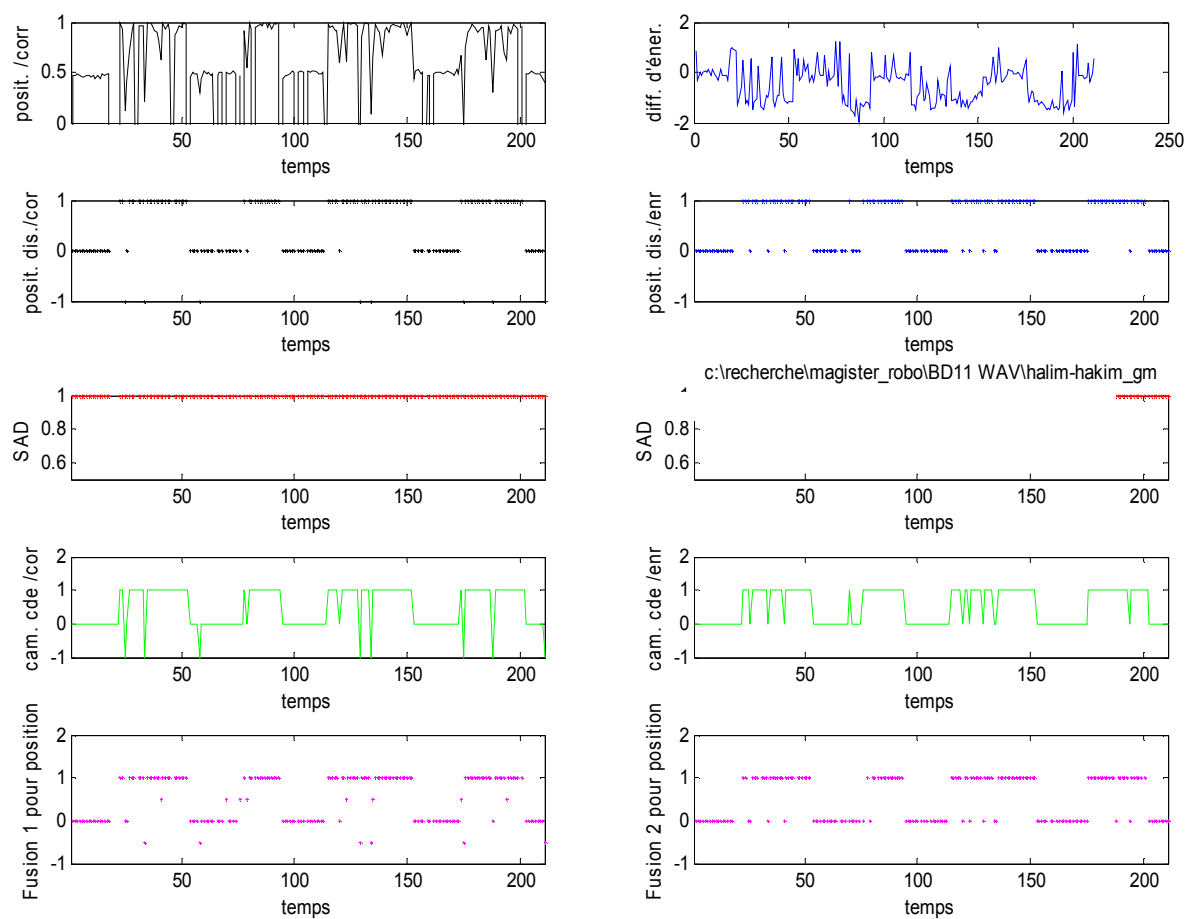


Fig 4.3 : Génération du signal de commande (en vert), à partir du signal de détection (obtenu par la méthode FCM à gauche en noir, et la méthode EDM à droite en bleu).

A titre d'explication, dans ce scénario nous avons deux locuteurs assis respectivement à droite et au Milieu.

# CHAPITRE V

## RESULTATS

### V.1- Introduction

Dans ce chapitre, on va exposer les résultats des différentes méthodes de localisation et de commande appliquées sur plusieurs scénarios distincts.

### V.2- Procédure des expériences réalisées

#### V.2.1-Configuration de la base de données utilisée

Notre base de données BD11 constituée de 6 scénarios, contient des locuteurs (masculins et féminins) avec des paroles naturelles assis dans des positions fixes sur lesquels on a réalisé des expériences de localisation de leurs positions.

#### V.2.2- Acquisition des signaux audio

Les signaux audio sont enregistrés par deux microphones (gauche et droite), et amplifiés par un pré-ampli.

#### V.2.3-Pré-traitement audio

L'écoute est le premier critère qui permet d'apprécier la qualité d'un signal, et de détecter les problèmes créés par le système de prise de son (Akchiche, 2007).

- L'écoute du signal de parole se fait à l'aide de logiciel **Gold Wave** qui manipule les fichiers aux formats audio.

#### V.2.4- Traitement audio

Pour indiquer la position de locuteur qui parle, on a implanté certains programmes de localisation spatiale en Matlab, en utilisant deux techniques de détection:

✚ Méthode du différentiel d'énergie (on en anglais Energy Differential Method **EDM**).

✚ Méthode de Generalized cross-correlation phase transform (**GCC –PHAT**).

Afin d'obtenir plus de précision dans la localisation de la position, on a décomposé le signal de parole en segments (fenêtres) de taille **0.5 secondes**. Les trois positions à détecter sont: la gauche (-1), le milieu (0) et la droite (1).

La fréquence d'échantillonnage des signaux est de 16 KHz.

### a) Méthode du différentiel d'énergie EDM

#### b.2) Résultats des expériences réalisées

En appliquant l'algorithme précédent aux différents scénarios à position fixe, on a obtenu les résultats illustrés dans les figures ci-après.

Ces résultats nous permettent de visualiser la correspondance de la distance  $d$  (entre le locuteur qui parle et l'un des microphones) avec la position.

Ainsi on peut faire une comparaison graphique entre les positions estimées expérimentalement (en bleu) et les positions de référence (en vert) et l'intersection des deux valeurs représente le taux de précision (partie hachurée en noir).

#### Scénario 1: Halim-Hakim\_ (droite-milieu)

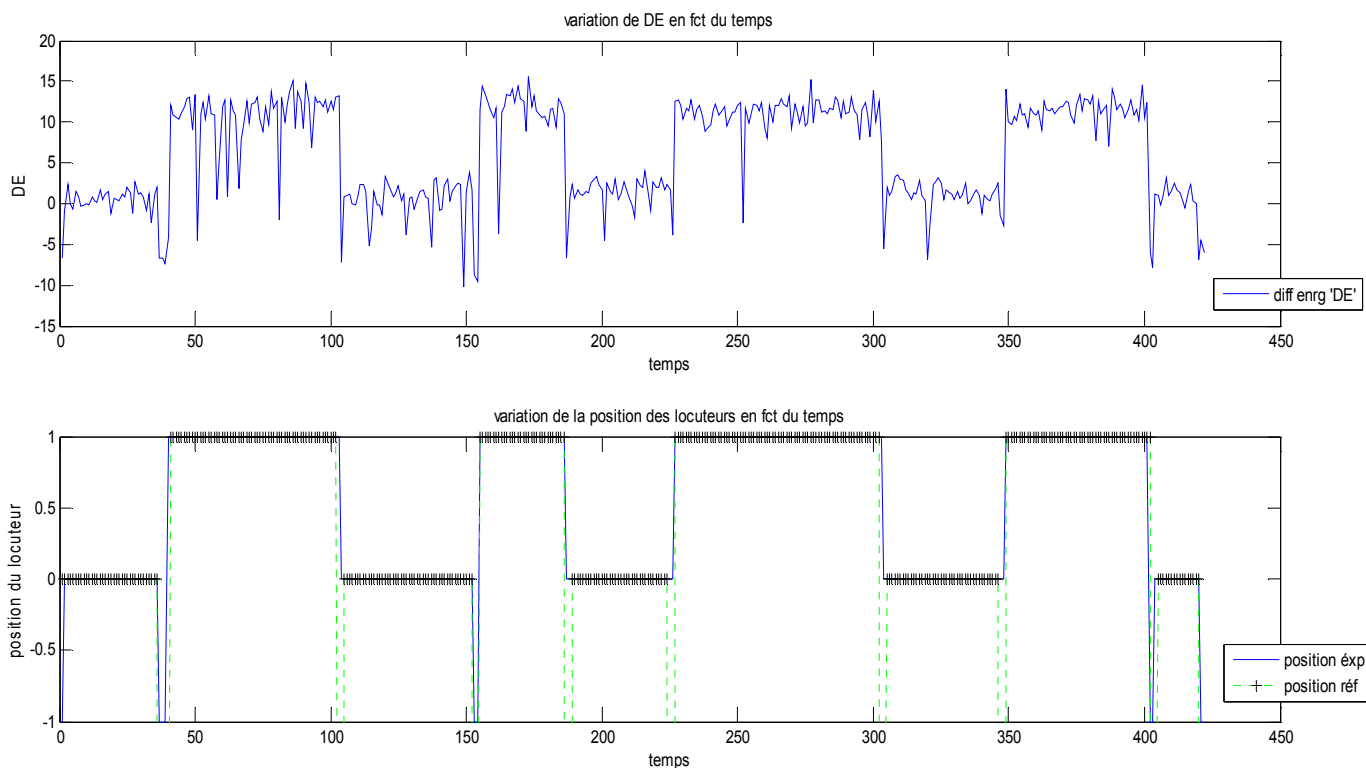


Figure 5.1 : Résultats de localisation obtenus par la méthode EDM.



## b) Méthode du Generalized cross-correlation phase transform (GCC –PHAT)

Pour la méthode de la **GCC-PHAT** : on calcule le retard ( $\tau$ ) entre les deux signaux qui correspond à la position du maximum et avec lequel on va déterminer la distance entre le locuteur qui parle et l'une des deux positions (gauche et droite).

Cette distance sera comparée à un seuil de détection bien choisi expérimentalement.

Voici des exemples de la **GCC-PHAT** sur des fenêtres d'un signal de parole:

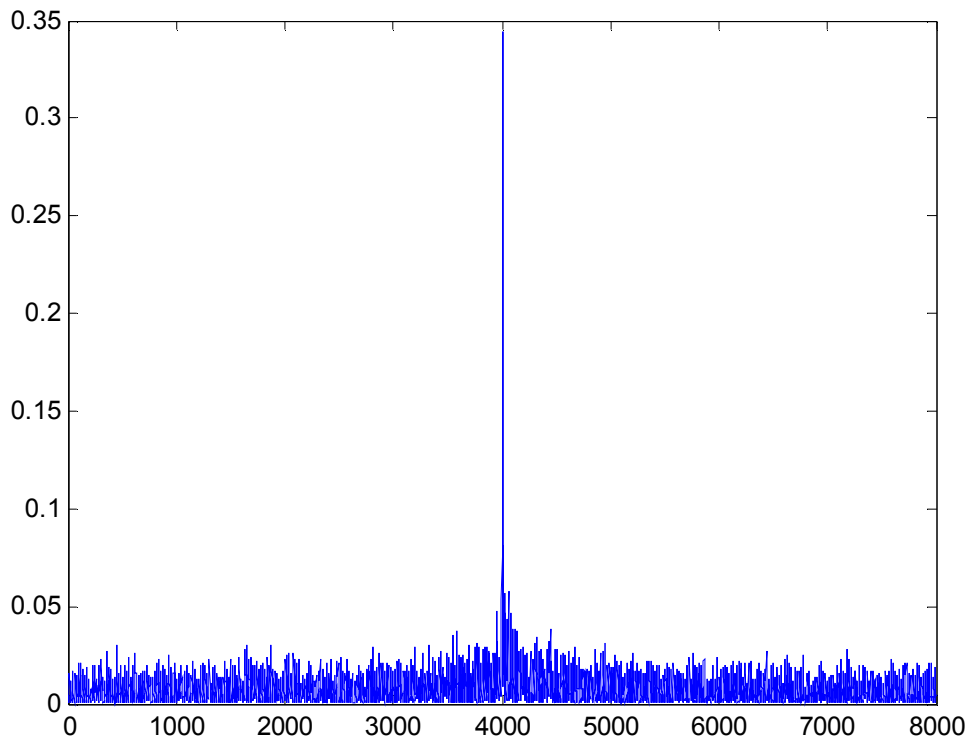


Figure 5.3 : exemple de localisation d'un locuteur au milieu (scénario 1)

- Le pic donne un retard  $\tau = 2,5 \text{ e}^{-4} \text{ s}$   
La distance **d** est :  $d=0,5426 \text{ m}$  ( $0,4 < d < 0,6$ )

Ainsi dans cet exemple : le locuteur qui parle est au milieu.

## c.2) Résultats des expériences réalisées

En appliquant l'algorithme précédent aux différents scénarios à position fixe, on a obtenu les résultats illustrés dans les figures ci-après.

Ces résultats nous permettent de visualiser la correspondance de la distance  $d$  (entre le locuteur qui parle et l'un des microphones) avec la position.

Ainsi, on peut faire une comparaison graphique entre les positions estimées expérimentalement (en rouge) et les positions de référence (en vert) et l'intersection des deux valeurs représente le taux de précision (partie hachurée en noir).

### Scénario 1: Halim-Hakim\_ (droite-milieu)

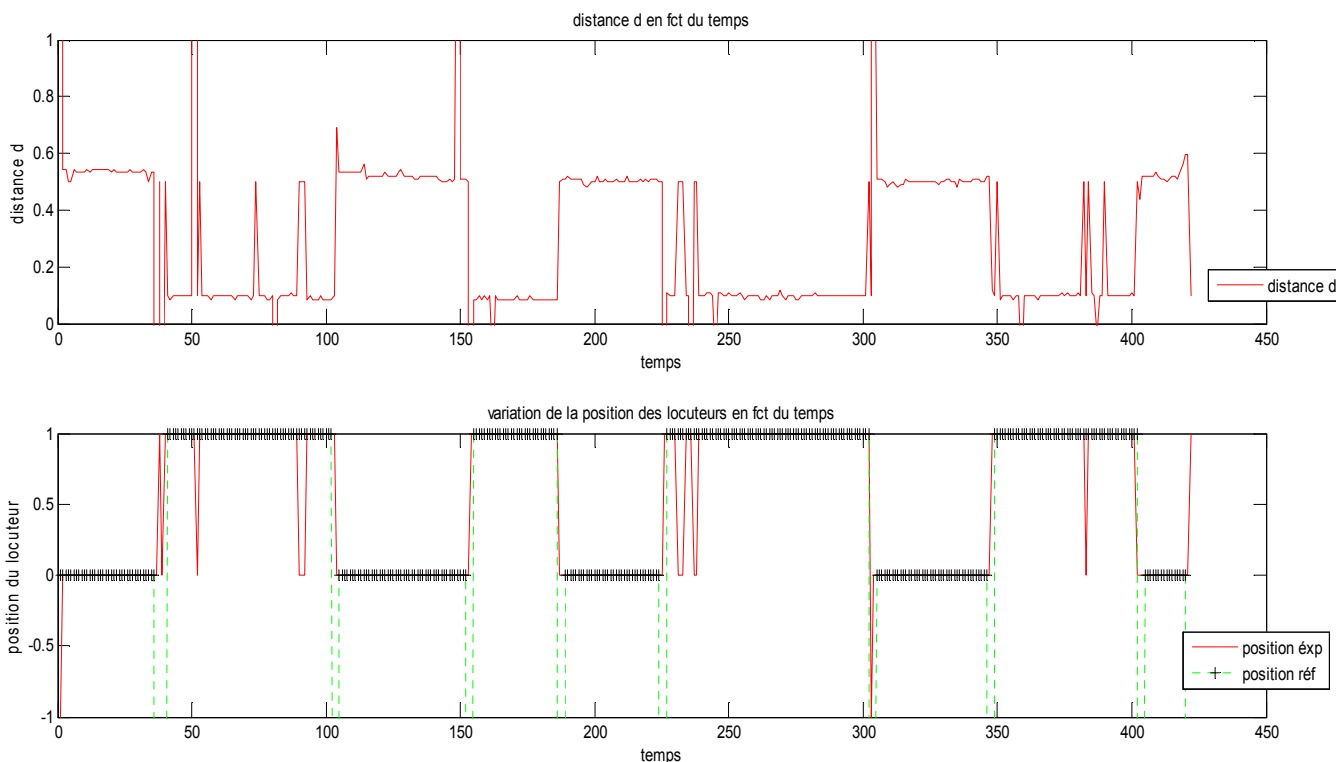


Figure 5.3 : Résultats graphiques obtenus par la méthode de la GCC-PHAT.

## Performances des résultats obtenus

Les différentes expériences de détection ont donné les résultats illustrés sur le tableau ci-dessous:

Tableau 5.1: Taux de précision de localisation en % par scénario.

techniques	EDM	GCC-PHAT
scénarios		
Scénario 1	99,5	97,01
Scénario 2	97,69	99,13
Scénario 3	90 ,46	99,13
Scénario 4	98,20	94,91
Scénario 5	91,78	97,6
Scénario 6	98,87	97,74

D'une manière graphique, nous obtenons la représentation illustrée sur la figure suivante :

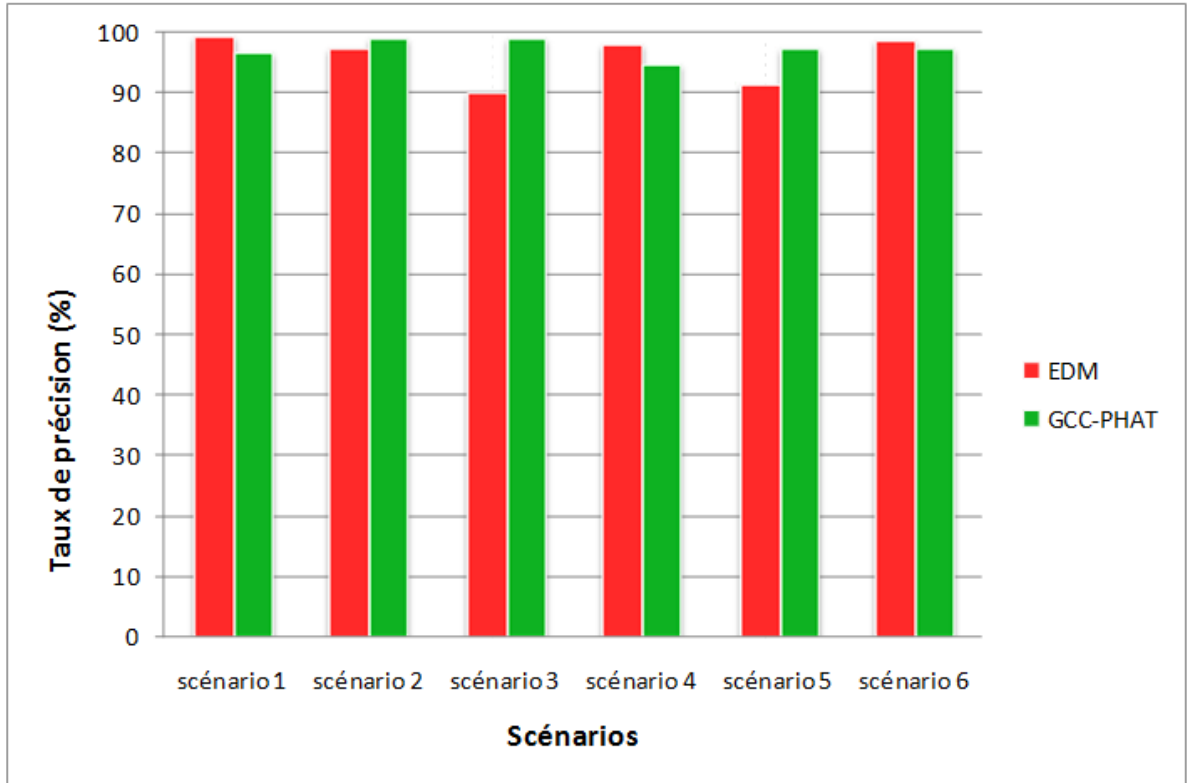


Figure 5.4: Taux de précision de localisation en % par scénario.

Pour pouvoir comparer les performances des deux méthodes, il est préférable de calculer la moyenne des différents taux de précision.

Les résultats obtenus sont présentés dans le tableau ci-dessous :

Tableau V.2: Performances moyennes.

	EDM	GCC-PHAT
Taux de localisation en %	96,03 %	97,59 %

### V.3 Expériences d'orientation de caméra

Dans cette série d'expériences nous avons procédé à la mise en fonction de notre système

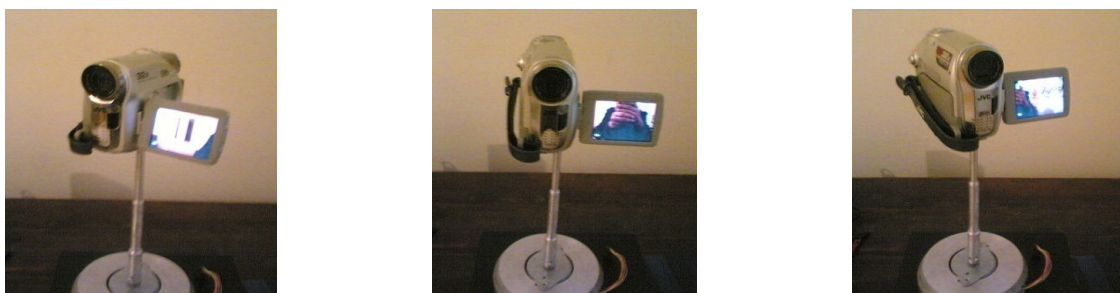
d'orientation de caméra. On a prévue pour cela deux types de systèmes d'orientation : un système de commande par moteur pas à pas et un autre système basé sur la commande virtuelle en utilisant la nouvelle technique VROI précédemment décrite.

### V.3.1 Expériences d'orientation de caméras par moteur pas à pas

Deux expériences d'orientation de caméra par moteurs pas à pas unipolaires ont été menées.

Une caméra vidéo numérique de type Sony, JVC et une caméra Webcam de type Kinstone ont été utilisées.

Avant d'entamer les tests expérimentaux, on a procédé aux différents tests d'orientation des deux caméras vers les trois positions fixes : gauche, milieu et droite. Ces tests ont montré une orientation correcte vers les différentes positions (voir figure suivante).



\_\_\_\_\_ Caméra dirigée vers : \_\_\_\_\_

*a) la droite*

*b) le milieu*

*c) la gauche*

Fig. 5.5 : Photos d'orientations de la caméra (JVC) placée sur un support motorisé. La caméra est orientée vers : la droite, le milieu et la gauche respectivement.

#### A. Expérience avec la caméra vidéo numérique JVC :

##### Avantages

Les avantages observés peuvent être résumés par les points suivants :

- Qualité d'image excellente ;
- Les orientations vers les positions : gauche, droite et milieu se font correctement ;
- Temps de réponse moyennement faible.

##### Inconvénients

Les inconvénients observés se résument par les points suivants :

- Oscillations de la caméra pendant le mouvement ;
- Bruit de claquement assez gênant dû au moteur pas à pas ;
- Inertie remarquable pour les déplacements très vifs.

### V.3.2 Expériences d'orientation de caméra utilisant la commande Virtuelle

L'utilisation de la caméra mobile rencontre beaucoup de problèmes qui sont dus principalement aux bruits du moteur, aux oscillations mécaniques, aux temps de réponse important et à l'importante inertie.

Ces raisons nous ont poussés à conduire une deuxième expérience qui consiste à assurer le suivi vidéo du locuteur par une caméra fixe. On a donc utilisé une nouvelle technique basée sur la région d'intérêt « Region Of Interest » qu'on a appelé : technique VROI.

Un exemple de capture par la nouvelle technique VROI est illustré dans la figure suivante :

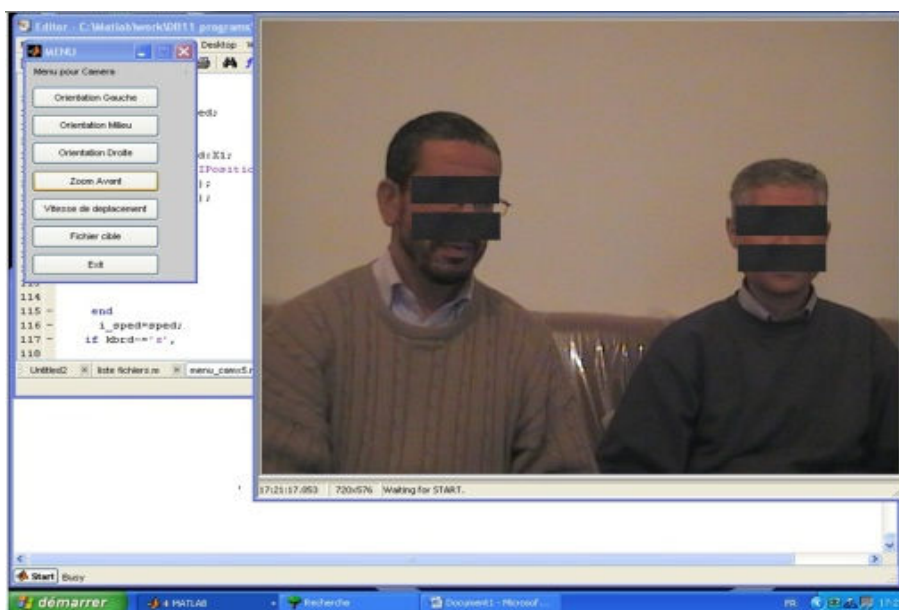


Fig. 5.6: Vidéo obtenue avec la caméra (JVC) en utilisant la technique VROI (zoom avant).

## PUBLICATIONS ASSOCIEES AU PROJET

### **-CHAPITRES DE LIVRE EDITES EN 2012/2013 :**

Nom du Livre : Advancing the Next-Generation of Mobile Computing: Emerging Technologies.

Titre: Automatic Speaker Localization and Tracking: Using a Fusion of the Filtered Correlation with the Energy Differential.

Site Web : <http://www.igi-global.com/book/advancing-next-generation-mobile-computing/58266>

### **PUBLICATIONS DANS DES REVUES INTERNATIONALES : 3 revues internationales**

Lien: <http://sayoud.net/Publications.html>

#### **1ere Revue internationale**

Journal: International Journal of Speech Technology - Springer.

Title: A New Approach of Speaker Clustering based on the Stereophonic Differential Energy.

Authors: S. Ouamour, H. Sayoud

References: Springer, Int. Jo. of Speech Techn. 2013; DOI: 10.1007/s10772-013-9199-z.

Website: <http://www.springer.com/engineering/signals/journal/10772>

#### **2ème Revue internationale**

Journal: International Journal of Speech Technology

Publisher: Springer Verlag

Indexed by: INSPEC (IET)

Title : A pertinent learning machine input feature for speaker discrimination by voice

Received: 31 August 2011 / Accepted: 23 January 2012. pp 181-190. Vol 15, no 02.

© Springer Science+Business Media, LLC 2012

Authors: S. Ouamour · H. Sayoud

References: Received: 31 August 2011 / Accepted: 23 January 2012

Int J Speech Technol (2012) 15:181–190

Website : <http://www.springer.com/engineering/signals/journal/10772>

#### **3ème Revue internationale**

Journal: International Journal of Networking and Virtual Organisations,

Publisher: Indersciences, Switzerland

Indexed by: INSPEC (IET)

Title : Virtual Speaker Tracking by Camera Using a Sound Source Localization with two Microphones

Authors: H. Sayoud, S. Khennouf and S. Ouamour

References: Vol. 12, No. 2, 2013. pp: 85-110.

Website : <http://www.inderscience.com/browse/index.php?journalID=66&action=coming>

Copyright © 2013 Inderscience Enterprises Ltd.

## CONFERENCES INTERNATIONALES

Lien : <http://sayoud.net/publications.html>

### 1- ICTA-2013 conference– Hammamet, Tunisia, 24-26 October, 2013.

Automatic Speaker Localization based on Speaker Identification -A Smart Room Application-  
<http://www.icta.rnu.tn/index.php?picta=1>

### 2- STA-2012 conference - Monastir, Tunisia, December 17-19, 2012.

Automatic Speaker Detection Using Support Vector Machines and an Interlaced Segmentation -Application on Telephonic Talks-. Accepted in the 11th International conference on Sciences and Techniques of Automatic control & computer engineering. Authors: Sayoud et al. <http://www.sta-tn.com> , pp 1517-1525.

### 3- ICCSE 2012 conference - London, UK, July 4-6, 2012,

Virtual System of Speaker Tracking by Camera Using an Audio-Based Source Localization. Conference ICCSE, July 4-6, 2012, London, UK. pp 819-822.

### 4- WCE 2011 conference - London, U.K., 6-8 July, 2011.

SPEAKER CLUSTERING OF STEREOPHONIC SPEECH SIGNAL USING SPATIAL AND SEQUENTIAL GATHERING. World Congress on Engineering 2011. Pp 1206-1210

### 5- WOSSPA 2011 conference - Tipaza May 9-11 2011.

Speaker localization using stereo-based sound source localization. In Systems, Signal Processing and their Applications (WOSSPA), 2011 7th International Workshop on (may 2011), pp. 231-234.

## -BASES DE DONNEES AUDIOVISUELLES FABRIQUEES (SUR DVD) EN 2013 :

Nous l'avons appelé ASDB2, les détails peuvent être retrouvés sur le site web : <http://sayoud.net/ASDB-2.pdf>

## -ENCADREMENT D'ETUDIANTS EN PFE et Doctorat:

Plusieurs étudiants de Master ont été encadrés et formés en 2012 et 2013 : 3 thèses de Master ont été soutenues (voir les détails plus bas).

-Encadrement d'un Master (soutenu en 2013) dans le thème du projet (Université de Msila). Titre : Réalisation d'un Système Automatique pour la Commande d'une Camera vers des Locuteurs Actifs. Celui-ci a soutenu en 2013. Encadreurs : H. Sayoud et S. Khennouf. Thésard : Aiche abderrazak.

-Encadrement de 2 étudiantes en Master (Master soutenu en 2012) dans le thème du projet (Université de Medea). Titre : Détection des Locuteurs par des Signaux Stéréophoniques. Celles-ci ont soutenu en 2012. Encadreurs : S. Ouamour et H. Sayoud. Thésards : Djemai Lynda et Sadi Hadjer.

- Encadrement d'un étudiant en Master. Titre : Suivi de séquence vidéo à l'aide de contours actifs, FEI, USTHB, Juin 2011. Encadreur : Prof A. Houacine.

-Encadrement d'une enseignante du partenaire économique (ISMAS) en Doctorat (Mlle Kennai Wassila). Encadreurs : S. Ouamour et M. Guerti.

## CONCLUSION

Dans ce travail de recherche, nous avons proposé un système automatique basé sur une seule caméra mobile à orientation autonome vers une cible sonore. Notre application était destinée à la supervision automatique des enregistrements audiovisuels dans les meeting-rooms, ce qui nous a conduits à orienter nos expériences vers l'utilisation des locuteurs fixes ou mobiles comme cibles sonores.

Ainsi plusieurs locuteurs assis devant une table, prennent la parole à tour de rôle, d'une manière naturelle, et une caméra mobile doit suivre automatiquement le locuteur parlant. Deux types d'orientations sont envisagés, et différentes techniques de suivi de locuteur ont été étudiées.

Pour ce faire, nous avons proposé trois méthodes de localisation du locuteur actif et deux systèmes de commande pour l'orientation de la caméra mobile vers cette cible sonore. La première méthode de localisation est basée sur une nouvelle technique de corrélation filtrée : nous l'avons appelée *Méthode FCM*. La deuxième méthode est basée sur le différentiel d'énergie : nous l'avons appelée *Méthode EDM*. La troisième méthode (existant déjà) est connue sous le nom de la GCC-PHAT.

Pour le système de commande, la première méthode utilise la commande par moteur pas-à-pas via un système mécatronique. La deuxième méthode est basée sur la nouvelle technique de commande virtuelle de caméra, par le biais du contrôle de la ROI (*Region Of Interest*) : nous l'avons appelée *Méthode VROI*. Plusieurs logiciels de commande virtuelle à base de VROI ont été réalisés durant ce projet PNR. Par ailleurs cette technique a fait l'objet d'une publication dans une revue internationale.

Durant ce travail de PNR, nous avons aussi fabriqué une base de données audio pour pouvoir faire les expériences de localisation nécessaires. On l'a appelée ASDB2 (voir site web : <http://sayoud.net/ASDB-2.pdf>).

Comme fruit de ce projet de recherche, on peut dire que l'on a bénéficié de beaucoup d'avantages dans notre laboratoire : nous avons pu encadrer et former plusieurs thésards, nous avons acquis un matériel qui nous a permis d'avancer et nous avons été motivé à travailler en équipe.



Concernant les publications liées au thème de ce projet, durant ces 2 dernières années, on a pu avoir :

- trois publications dans des revues internationales ;
- un chapitre de livre édité ;
- et cinq conférences internationales.

Finalement, et pour être honnête il faut avouer que ce que l'on a fait est relativement peu devant ce qu'on pourrait vraiment réaliser si on avait travaillé plus sérieusement encore.

## REFERENCES BIBLIOGRAPHIQUES

- [Argentieri, 2005] S. Argentieri, P. Danes, P. Soueres, " Conception d'un système de localisation de source sonore large bande par formation de voie pour des applications robotiques ", 2005. GRETSI - Actes de Colloques, 2005.
- [Argentieri, 2006] S. Argentieri " Conception d'un capteur sonore pour la localisation de source en robotique mobile ", thèse de doctorat. 8 Décembre 2006, Université Paul Sabatier, Toulouse, France.
- [Asano, 1999] F. Asano, H. Asoh & T. Matsui " Sound Source Localization and Signal Separation for Office Robot Jijo-2 ", IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems, pages 243–248, Aug. 1999.
- [Bechler, 2003] D. Bechler, M. Grimm, K. Kroschel " Speaker tracking with a microphone array using Kalman filtering ", Advances in Radio Science 2003 pp : 113–117 Institut für Nachrichtentechnik, Universität Karlsruhe, Germany.
- [Bellanger, 1987] M. Bellanger " Traitement Numérique du Signal", Collection technique et scientifique des télécommunications, édition MASSON, Paris 1987.
- [Boite, 1987] R. Boite, M. Kunt " Traitement de la Parole ", complément traité d'électricité. Presses Polytechniques Romandes, Lausanne, Paris, 1987.
- [Brooks 1999] R. Brooks, T.B. Senior & P.L.E. Uslenghi. The Cog Project : Building a Humanoid Robot. C.L. Nehaniv, editeur, Computations for Metaphors, Analogy and Agents. Springer Verlag, 1999.
- [Brown, 1993] G. J. Brown " Computational Auditory Scene Analysis : A representational approach ". Journal of the Acoustical Society of America, vol. 94, no 4, page 2454, October 1993.
- [Bruel, 2003] Bruel and Kjør " Weatherproof microphone unit type 4184 ", Product Data, 2003.
- [Cavaco, 1999] S. Cavaco & J. Hallam " A Biologically Plausible Acoustic Azimuth Estimation System ". Third International Workshop on Computational Auditory Scene Analysis of the Sixteenth International Joint Conference on Artificial Intelligence, pages 78–87, 1999.
- [Cox, 1987] H. Cox et al. " Robust adaptive beamforming ", IEEE Trans. on Acoustics, Speech and Signal Proc., Octobre 1987.

- [Coulon, 1987] F. Coulon " Théorie et Traitement du Signal ", Presses Polytechniques Romandes, Lausanne, 1987.
- [Desclaux, 1997 ] J.F Desclaux et al. " Electronique Pratique, CDROM ", 1997.
- [Djekouane, 1997] A. O. Djekouane " Localisation et Guidage par Caméra Embarquée d'un Robot Mobile ". Thèse de Magister, USTHB, Alger, 1997.
- [Hartmann, 1999] W. M. Hartmann. " How We Localize Sound ", Physics Today, vol. 52, no 11, pages 24–29, 1999.
- [Hebrank, 1974] J. Hebrank & D. Wright " Spectral Cues Used in The Localization of Sound Sources on The Median Plane ". Journal of the Acoustical Society of America, vol. 56, no 6, pages 1829–1834, December 1974.
- [Huang, 1995] J. Huang, N. Ohnishi & N. Sugie " A Biomimetic System for Localization and Separation of Multiple Sound Sources ". IEEE Transactions on Instrumentation and Measurement, vol. 44, no 3, pages 733–738, June 1995.
- [Huang, 1997] J. Huang, N. Ohnishi & N. Sugie " Separation of multiple sound sources by using directional information of sound source ", vol. 1, no 4. Springer Tokyo, 1997.
- [Kim, 2008] U. Kim, J. Kim, D. Kim, H. Kim and B. You, " Speaker Localization on a Humanoid Robot's Head using the TDOA-based Feature Matrix ", IEEE International Symposium on Robot and Human Interactive Communication RO-MAN 2008.
- [Kirkwood, 2003] B. C. Kirkwood, " Acoustic Source Localization Using Time-Delay Estimation ", M.Sc. thesis. August 4, 2003, Technical University of Denmark.
- [Kumon, 2005] M. Kumon, T. Shimoda, R. Kohzawa & Z. Iwai " Audio Servo for Robotic Systems with Pinnae ", IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 885–890, Aug. 2005.
- [Kunt, 1996] M. Kunt " Traitement numérique des signaux ". Edition technique et documentation, Lavoisier, 3ème édition, Lausanne, Suisse, 1996.
- [Lathoud, 2006] G. Lathoud " Spatio-temporal analysis of spontaneous speech with microphone arrays ", PhD thesis. December 22, 2006. Ecole Polytechnique Fédérale de Lausanne, Switzerland.
- [Lauchle, 1984] G. C. Lauchle, " Effect of turbulent boundary layer flow on measurement of acoustic pressure and intensity ". Noise Control Engineering Journal, 23(2): 52–59, 1984.

- [Lee, 1995] H.S. Lee, A.C Tsoi " Application of Multi-Layer Perceptron in Estimating Speech/Noise Characteristics for Speech Recognition in Noisy Environment ". Speech Communication, Vol. 17, Number1-2, pp. 59-76, 1995.
- [Litovsky, 1999] R. Y. Litovsky, H. S. Colburn, W. A. Yost & S. J. Guzman " The Precedence Effect ". Journal of the Acoustical Society of America, vol. 4, pages 1633–1654, October 1999. Structuring of Lectures, Proc. Of ACM Multimedia'99, Orlando.
- [Liu, 2000] Q. Liu, Y. Rui, A .Gupta, and J.J. Cadiz " Automating Caméra Management for Lecture Room Environments ". Technical Report No: MSR-TR-2000-90, Microsoft Research, September 2000. Brazil, May 6-9, 2008, pp. 527-530.
- [Lopez-Poveda, 1996] E. A. Lopez-Poveda & R. Meddis " A physical model of sound diffraction and reflections in the human concha ". Journal of the Acoustical Society of America, vol. 100, no 5, pages 3248–3259, November 1996.
- [Maganti, 2006] H. K. Maganti and D. G. Perez, " Speaker Localization for Microphone ArrayBased ASR : The Effects of Accuracy on Overlapping Speech ", 13ICMI'06, Novembre 2–4, 2006, Banff, Alberta, Canada.
- [Marchand, 2008] E. Marchand, " Positionnement relatif d'une caméra et d'une source lumineuse en utilisant les gradients d'intensité de l'image" 16<sup>e</sup> congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle, RFIA'08 Bretagne Atlantique, Rennes, France.
- [Marchildon, 2004] <http://nicolas.marchildon.net/electro/pap/index.html>
- [Matlab, 1992] Matlab High-Performance Numeric Computation & Visualization Software, The Math Works Inc, Reference Guide, August 1992.
- [Matusaka, 1999] Y. Matusaka, T. Tojo, S. Kubota, K. Furukawa, D. Tamiya, K. Hayata, Y. Nakano & T. Kobayashi " Multiperson Conversation via Multi-modal Interface - A Robot who Communicate with Multi-user - ". Eurospeech, pages 1723–1726, 1999.
- [Mennen, 2008] I. Mennen, F. Schaeffler, and G. Docherty, " A methodological study into the linguistic dimensions of pitch range differences between German and English ", Speech Prosody Conference, Campinas,
- [Mukhopadhyay, 1999] S. Mukhopadhyay, B. Smith " Passive Capture and
- [Mumolo, 2002] E. Mumolo, M. Nolich, and G. Vercelli, " Algorithms for acoustic localization based on microphone array in service robotics ", Robotics and

Autonomous Systems, 42:69–88, 2002.

- [Murray, 2005] J. Murray, S. Wermter & H. Erwin " Auditory Robotic Tracking of Sound Sources using Hybrid Cross-Correlation and Recurrent Networks ". IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 891–896, 2005.
- [Nakadai, 2000a] K. Nakadai, T. Lourens, H. G. Okuno & H. Kitano " Active Audition for Humanoid ". 17<sup>th</sup> National Conference on Artificial Intelligence, 2000.
- [Nakadai, 2000b] K. Nakadai, T. Matsui, H.G. Okuno & H. Kitano " Active Audition System And Humanoid Exterior Design ". International Conference on Intelligent Robots and Systems, vol. 2, pages 1453–1461, Nov. 2000.
- [Okuyama, 2002] F. Okuyama, J. Y. Takayama, S. Ohya & A. Kobayashi " A study on determination of a sound wave propagation direction for tracing a sound source ". Proceedings of the 41st SICE Annual Conference, vol. 2, pages 1102–1104, Aug. 2002.
- [Ouamour, 2009<sup>1</sup>] S. Ouamour, H. Sayoud & M. Guerti, " Optimal Spectral Resolution in Speaker Authentication, Application in noisy environment and Telephony ". International Journal of Mobile Computing and Multimedia Communications (IJMCMC). Avril-Juin 2009. Page to page: 36-47.
- [Ouamour, 2009<sup>2</sup>] S. Ouamour, H. Sayoud, " Speaker Discrimination on Broadcast News and Telephonic Calls Using a Fusion of Neural and Statistical Classifiers ". The Mediterranean Journal of Computers and Networks (MedJCN), 2009.
- [Strybel, 1984] T. Z. Strybel & D. R. Perrott " Discrimination of Relative Distance in the Auditory Modality : the Success and Failure of the Loudness Discrimination Hypothesis ". Journal of the Acoustical Society of America, vol. 76, no 20, page 318, 1984.
- [Parker, 2001] ParkerVision, <http://www.polycom.com>
- [Purves, 2004] D. Purves, G. J. Augustine, D. Fitzpatrick, W. C. Hall, A.S. LaMantia, J. O. McNamara & S. M. Williams. " Neuroscience ". Sinauer Associates, 3<sup>ème</sup> édition, 2004.
- [Raichel, 2000] D. R. Raichel, " The Science and Applications of Acoustics ". AIP Series in Modern Acoustics and Signal Processing. Springer-Verlag, New York, 2000.
- [Rasmussen, 1997] K. Rasmussen " Calculation methods for the physical properties of air used in the calibration of microphones ". Department of Acoustic Technology, Technical University of Denmark, Lyngby, Denmark. Report Number PL-11b, 1997.
- [Sayoud, 2003] H. Sayoud et al., 'ASTRA' "An Automatic Speaker Tracking System

Based on SOSM Measures and an Inter-laced Indexation". Acta Acoustica, International Journal of Acoustics. Issue : No 4, Vol 89, Année : 2003, Page-to-page : 702-710.

- [Sayoud, 2003<sup>T</sup>] H. Sayoud, Thèse de Doctorat d'état. " Reconnaissance Automatique du locuteur : Approche connexionniste ". Novembre 2003, USTHB Alger.
- [Schmidt, 1979] R. O. Schmidt. " Multiple Emitter Location and Signal Parameter Estimation ". RADC Spectrum Estimation Workshop, Octobre 1979.
- [Sun, 2005] X. Sun, J. Foote, D. Kimber, B. S. Manjunath, " Region of interest extraction and virtual caméra control based on panoramic video capturing ". Dept. of Electr. & Comput. Eng., Univ. of California, USA, Multimedia, IEEE Transactions Octobre 2005 page(s): 981- 990.
- [Tarel, 1994] J-P. Tarel, " Calibration de Caméra Fondée sur les Ellipses ". Rapport de Recherche N02200, I.N.R.I.A, France, Jan 1994.
- [Vermaak, 2001] J. Vermaak, M. Gangnet, A. Blake, P. Perez, " Sequential Monte Carlo fusion of sound and vision for speaker tracking ". Eighth IEEE International Conference, July 2001. Vol : 1, pp : 741-746. Vancouver, Canada.
- [Wang, 2004] Q. H. Wang, T. Ivanov & P. Aarabi " Acoustic robot navigation using distributed microphone arrays ". Information Fusion, vol. 5, no 2, pages 131–140, June 2004.
- [Zotkin, 2001] D. Zotkin, R. Duraiswami, H. Nanda, L.S. Davis " Multimodal tracking for smart video conferencing ". This paper appears in : Multimedia and Expo, 2001. ICME 2001. IEEE International Conference on Publication Date: 22-25 Août. 2001, pages: 36- 39.

## Annexe:

Dépenses effectuées à partir des 105 Millions de centimes octroyés par le PNR.

Nature des fournitures et travaux commandés (Matériels et instruments scientifiques et audiovisuels)	Quantité	Prix unitaire (DA)	Valeur (DA)	Observations
Carte d'acquisition audio professionnelle à 2 entrées. Type : MobilePre MK II	1	15 077,00	15 077,00	Matériel et instrument scientifique et audiovisuel
Microphone professionnel HI-FI Shure 58 Beta	2	21 451,00	42 902,00	
Camera haute définition sony HDR-CX210	1	43 793,00	43 793,00	
PC de bureau HP Pavillon P2 G630 RAM 2G/ HDD 500	1	45 406,00	45 406,00	
Ordinateur Portable Toshiba i7 15,6" L855-10W	3	117 200,00	351 600,00	
Ordinateur Portable HP i3 17,3" Probook 4730S	1	76 700,00	76 700,00	
Imprimante Laser Multifonction MX14	2	22 500,00	45 000,00	
Vidéo Projecteur SVGA Epson S11	2	53 125,00	106 250,00	
Total HT			726 728,00	
TVA (7%)				
TVA (17%)			123 543,76	
<b>Total TTC</b>			<b>850 271,76</b>	

Nature des fournitures et travaux commandés (Papeterie et fourniture de bureau )	Quantité	Prix unitaire (DA)	Valeur (DA)	Observations
RAME PAPIER EXTRA BLANC A4/80 GRS DOUBLE A	50	416,00	20 800,00	Papeterie et fourniture de bureau
RAME CHEMIS DOSSIER DIF, COULEUR 250F GALAXI	2	1 400,00	2 800,00	
RAME SOUS CHEMISE 500F	1	1 200,00	1 200,00	
MARQUEUR TABLEAU BLANC BLEU MW85TETE POINTUE PENDEL	10	67,00	670,00	
Total HT			25 470,00	
TVA (7%)				
TVA (17%)			4 329,90	
<b>Total TTC</b>			<b>29 799,90</b>	

Nature des fournitures et travaux commandés (Mobilier de laboratoire )	Quantité	Prix unitaire (DA)	Valeur (DA)	Observations
Bureau simple 1,20 -3 Tiroirs	1	5 800	5 800,00	Mobilier de laboratoire
Chaise tournante confortable	1	15 000	15 000,00	
Armoire 1m60 - 2 portes	1	15 000	15 000,00	
Table pour PC	2	2 000	4 000,00	
Total HT			39 800,00	
TVA (7%)				
TVA (17%)			6 766,00	
<b>Total TTC</b>			<b>46 566,00</b>	

Nature des fournitures et travaux commandés (Renouv. du matériel informatique, logiciel et consommable informatique )	Quantité	Prix unitaire (DA)	Valeur (DA)	Observations
Toner M1400/MX14	6	12 375,00	74 250,00	Logiciel et consommable informatique
Antivirus KESPERSKY pour 3 poste	4	2 564,00	10 256,00	
Total HT			84 506,00	
TVA (7%)				
TVA (17%)			14 366,02	
<b>Total TTC</b>			<b>98 872,02</b>	