République Algérienne Démocratique et Populaire
Ministère de l'enseignement Supérieur et de la Recherche Scientifique
Université de Dr Yahia Fares -MEDEA-

Faculté d'Electronique et d'Informatique

Domaine Sciences Techniques

Mémoire de Master

Génie Electrique

Option: Systèmes de Télécommunications

Thème

DETECTION SPATIALE DES LOCUTEURS PAR DES SIGNAUX STEREOPHONIQUES.

Proposé et dirigé par :

Mme OUAMOUR Siham.

Mr SAYOUD Halim

Présenté par :

■ **M^{elle} DJEMAI Lynda.**

■ **M^{elle} SADI Hadjer.**

JUILLET 2012

Remerciements

*D'abord nous remercierons **DIEU** pour la force que nous a donné afin de terminer ce travail dans la tranquillité.*

Nos vifs remerciements à nos familles pour leur soutien moral et physique durant toutes nos années d'étude ainsi que leur compréhension.

*Nous tiendrons à remercier notre promotrice **Mme. Siham OUAMOUR,** Maitre de Conférences à l'USTHB et notre promoteur **Mr. Halim SAYOUD,** Professeur à l'USTHB, pour nous avoir suivi, orienté et encouragé avec leur modestie et leur serviabilité.*

Une pensée particulière à tous nos enseignants du primaire à l'université pour nous avoir donné un bon niveau d'instruction.

Ainsi, nous présentons toutes nos gratitudes pour tous ceux qui nous ont aidés de près ou de loin par encouragement ou par contribution afin de pouvoir continuer surtout bien nos collègues et nos amis.

Nous remercions également le président et les membres de jury.

LYNDA

et

HADJER

TABLE DES MATIÈRES

Pages

Introduction générale……..…1

Chapitre I : Généralités sur la localisation de source sonore.

I.1- Introduction……...3

Ⅰ.2-C'est quoi un son ?.....3

Ⅰ.3-Physiologie de l'oreille humaine……4

Ⅰ.4-Mécanisme d'audition…….5

Ⅰ.5-La localisation de la source sonore……6

Ⅰ.5.1-Localisation à l'azimut (la direction)……...6

a) Les Différences de Niveau Interauriculaires (DNI)……..6
b) Les Différences de Temps Interauriculaires (DTI)……...8

Ⅰ.5.2-Localisation de la profondeur (l'élévation)……...……..9

Ⅰ.5.3-La perception de la distance de la source…….....10

a) En champ libre……10
b) Dans un espace réverbérant.....10

I.6- Conclusion……....11

Chapitre II : Méthodes de localisation des sources sonores.

II.1- Introduction……12

II.2- Techniques de localisation de la position……...12

II.2.1- Méthode du différentiel d'énergie (EDM)……12

a) Calcul du différentiel d'énergie……12

b) Indication de la position localisée……...13

II.2.2- Méthodes basées sur le calcul du retard……..14

II.2.2.1- Calcul de la distance entre le locuteur et les deux microphones……..14

II.2.2.2- Calcul du retard……..16

a) Méthode de la corrélation…….16

a.1) Définition…….16

a.2) L'autocorrélation « autocorrélation »……..……16

a.3) L'inter-corrélation « cross corrélation »……16

a.4) Estimation du retard par l'inter-corrélation « cross corrélation »……...17

b) Méthode de la corrélation filtrée (FCM)……18

c) Méthode de Generalized cross-correlation phase transform(GCC –PHAT)..19

c.1) Définition……19

c.2) Dérivation de la GCC……20

c.3) Phase Transform (PHAT)……..21

c.4) La GCC – PHAT……..22

II.3- Exemple d'application de localisation ……23

Cas d'un signal sinusoïdal (cas théorique)……..23

II.4- Conclusion……...26

Chapitre III : Base de données expérimentale.

III.1- Introduction……27

III.2-Expériences réalisées……27

III.3- Conception de la base de données DB11……...29

III.3.1- Conditions de réalisation……..29

III.3.2-Scénarios de la base de données BD11……...30

III.3.3- Configurations de la base de données BD11……....31

III.3.4-Discours enregistrés……31

III.4- Dispositifs matériels et logiciels utilisés……32

III.4.1- Schéma synoptique du système……32

III.4.2- Description des différents blocs du système……32

a) Bloc d'acquisition……32

a.1) Microphones……..33

a.2) Préampli audio……...33

b) Bloc de traitement et de commande……35

b.1) Unité de traitement……..35

- Pré-traitement audio avec le logiciel Gold Wave……...35
- Traitement audio avec le logiciel Matlab……....37

b.2) Circuit de puissance……37

c) Bloc d'exécution……37

III.5- Conclusion……..37

Chapitre IV : Expériences et résultats.

IV.1-Introduction……38

IV.2- Procédure des expériences réalisées……38

IV.2.1- Configuration de la base de données BD11……38

IV.2.2- Acquisition des signaux audio……..38

IV.2.3- Pré-traitement audio avec le logiciel Gold Wave.……38

IV.2.4- Traitement audio avec le logiciel Matlab……....40

a) Méthode du différentiel d'énergie EDM……..40

a.1) Algorithme…….....41

a.2) Résultats des expériences réalisées…….....41

b) Méthode de Generalized cross-correlation phase transform(GCC –PHAT)....48

b.1) Algorithme…….50

b.2) Résultats des expériences réalisées……50

IV.3- Statistiques des résultats obtenus……..54

IV.4- Conclusion……57

Conclusion générale……....58

Bibliographie……..59

TABLE DES FIGURES

Figure 1.1: *Ondulation du tympan en fonction du temps pour un son grave et un son aigu*...3

Figure 1.2: *Système auditif humain*……4

Figure 1.3**:***La localisation naturelle de l'espace sonore*……6

Figure 1.4: *La DNI en fonction de l'angle d'azimut pour différentes fréquences*……..8

Figure 1.5**:***Le cône de confusion*……9

Figure 1.6**:***La Fonction de Transfert Anatomique*……10

Figure 1.7:*Estimation de la distance d'une voix parlée ou chuchotée en espace libre*……11

Figure 2.1 : *Schéma représentant un locuteur entre deux microphones*…….....14

Figure 2.2 : *Estimation du retard par la méthode de* corrélation……17

Figure 2.3 : *Estimation du retard par la méthode de corrélation filtrée*…….18

Figure 2.4: *TDOA entre chaque deux* microphone.....19

Figure 2.5 : *Estimation du retard par la méthode de GCC*-PHAT……..22

Figure 2.6 : *Deux sinusoïdes légèrement décalées*……..24

Figure 2.7: *L'allure des différentes techniques pour le calcul du retard*……*25*

Figure 3.1: *Schéma synoptique de la configuration de la meeting-room*……27

Figure 3.2 : *Photos montrant des locuteurs dans des différentes positions*……30

Figure3.3: *Schéma synoptique de la configuration fixe*……*31*

Figure 3.4:*Schéma synoptique du système automatique réalisé*……32

Figure 3.5 ***:*** *Photo des microphones utilisés dans les enregistrements*……33

Figure 3.6: *Schéma électrique du Pré-ampli audio utilisé*……34

Figure 3.7: *Pré-ampli audio utilisé*……34

Figure 3.8: *Photo montrant le pré-traitement en utilisant le Gold Wave*……36

Figure 4.1: *Visualisation des signaux audio*……39

Figure 4.2 :*Résultats de localisation obtenus par la méthode EDM pour scénario 1*…….42

Figure 4.3 :*Résultats de localisation obtenus par la méthode EDM pour scénario 2*…….42

Figure 4.4 :*Résultats de localisation obtenus par la méthode EDM pour scénario 3*…….43

Figure 4.5 :*Résultats de localisation obtenus par la méthode EDM pour scénario 4*……..43

Figure 4.6 :*Résultats de localisation obtenus par la méthode EDM pour scénario 5*……..44

Figure 4.7 :*Résultats de localisation obtenus par la méthode EDM pour scénario 6*……..44

Figure 4.8 *: exemple de localisation d'un locuteur au milieu (scénario 1)*……..45

Figure 4.9*: exemple de localisation d'un locuteur à droite (scénario 1.)*……46

Figure 4.10:*Résultats obtenus par la méthode de la GCC-PHAT pour scénario 1*……..48

Figure 4.11:*Résultats obtenus par la méthode de la GCC-PHAT pour scénario 2*……..48

Figure 4.12:*Résultats obtenus par la méthode de la GCC-PHAT pour scénario 3*……..49

Figure 4.13:*Résultats obtenus par la méthode de la GCC-PHAT pour scénario 4*……..49

Figure 4.14:*Résultats obtenus par la méthode de la GCC-PHAT pour scénario 5*……..50

Figure 4.15:*Résultats obtenus par la méthode de la GCC-PHAT pour scénario 6*……..50

Figure 4.16: *Taux de précision de localisation de la position par scénario*……..52

Figure 4.17: *La moyenne des taux de précision par méthode*……..*53*

LISTE DES TABLEAUX

Tableau 2.1: *Résultats de simulation des différentes méthodes*……25

Tableau 3 .1 : *Récapitulatif des bases de données réalisées*……...28

Tableau 3 .2 : *Résultats de localisation des locuteurs*……28

Tableau 4.1: *Taux de précision de localisation de la position par scénario*……..51

Tableau 4.2: *La moyenne des taux de précision par méthode*……..53

Au cours de ces dernières années, le problème de la localisation d'une source sonore dans l'espace a connu un intérêt croissant par la communauté scientifique qui veut rapprocher de plus en plus de la capacité humaine dans la détection du son (Brutti, 2008).

Plusieurs applications de traitement audio peuvent obtenir des avantages substantiels sur la connaissance de la position de la source qui émet le signal sous processus et dans la plupart de ces applications, on estime cet emplacement. Ces estimations doivent être produites à un rythme élevé et avec une latence très faible, mais ils ont besoin d'être précis.

Pour cette raison, de grands efforts ont été consacrés à enquêter sur ce domaine de recherche et plusieurs approches ont été proposées ce qui a enrichi le domaine de l'intelligence artificielle de point de vue conceptions de matériel ainsi que la mise en œuvre de logiciels temps-réel (Hoang, 2009).

En général, il existe deux catégories d'algorithmes qui ont été explorées pour les systèmes réels: les algorithmes à deux étages et les algorithmes à un étage. L'exécution et le coût de calcul de ces deux catégories varient selon les conditions des environnements, les données étant traitées, les moyens mis en œuvre ces algorithmes et aussi les causeurs eux-mêmes.

- L'algorithme à deux étages se fait en deux étapes :

Dans la première étape, ou l'étape de l'estimation du temps de retard (TDE: Time Delay Estimation), il se produit un temps de retard de l'arrivée (TDOA: Time Difference Of Arrival) du son entre des paires de microphones.
Dans la deuxième étape, les courbes hyperboliques sont générées à partir des estimations disponibles de temporisation et de données de position du microphone, et puis on doit estimer l'emplacement de la source de l'intersection de ces courbes dans un certain sens optimal.

- L'algorithme à un étage ne traite pas d'une manière paire, mais exploite la multitude de microphones, afin de surmonter la limitation indiquée par la prise de décisions rapide et de réverbération (Hoang,2009).

Traditionnellement, les algorithmes de la localisation des sources sonores sont fondés sur une estimation de la différence du temps d'arrivée (TDOA) aux deux paires de microphone à partir du quel on peut déduire des informations sur la position spatiale d'un émetteur source (Brutti, 2008).
Ce type de système basé sur des réseaux de microphones a été utilisé dans de nombreuses applications, telles que la téléconférence, la télésurveillance, la reconnaissance de la parole, la caractérisation du causeur et la capture de la voix dans des environnements réverbérants (Hoang,2009).

C'est pourquoi le but de notre projet est de proposer un système permettant de détecter la position d'un locuteur donné en utilisant son signal de parole enregistré par deux microphones placés en opposition et séparés par une distance fixe .

Généralement, dans les meeting-rooms, les interlocuteurs sont assis autour d'une table à des positions fixes. Ceux-ci se partagent une discussion assez continue et hiérarchique. On utilisera 2 microphones pour la capture du signal de parole. A partir de ce signal, un traitement adéquat permettra de trouver lequel des locuteurs a parlé et où il se trouve.

Plan de travail :

Nous avons organisé notre thèse en quatre chapitres :

- Le premier chapitre porte sur une recherche bibliographique sur le système auditif humain ainsi que le domaine de la localisation spatiale de source sonore.

- Le second chapitre expose certaines techniques de localisation spatiale de la source sonore telles que: la méthode de la corrélation, la méthode de la corrélation filtrée, la méthode à énergie différentielle et GCC-PHAT (Generalised Cross Correlation-PHAse Transform).

- Dans le troisième chapitre, on va discuter la base de données audio réalisée expérimentalement DB11.

- Le dernier chapitre est consacré à l'évaluation de nos algorithmes sur la base de données stéréophonique DB11 et l'interprétation des résultats obtenus.

Enfin, nous terminons par une conclusion générale, des perspectives et des références bibliographiques pouvant éventuellement éclaircir certains points.

Chapitre I

Généralités sur la localisation de source sonore

Chapitre II

Méthodes de localisation des sources sonores

Chapitre III
Base de données expérimentale

Chapitre IV
Expériences et résultats

Ⅰ.1- Introduction

Le système auditif humain constitue un ensemble très performant. En effet, avec seulement deuxrécepteurs (oreilles), le cerveau est capable, et très rapidement, de prendre en compte tous les sons dans notre entourage et de nous indiquer sur leurs positions d'émission.
De nombreuses équipes de recherches dans le monde entier se sont penchées sur l'idée de pouvoir, comme pour l'être humain, localiser la source d'émission d'un son (Margueron, 2000).

Leurs étudesmènent à plusieurs méthodes de localisation spatiale des sources sonores.
Pour y pouvoir étudier ces méthodes, il faut d'abord bien comprendre ce qu'est un son, puis comment on peut localiser leur source.

Ⅰ.2-C'est quoi un son ?

Les sons "purs" sont des oscillations régulières de la pression de l'air (les spécialistes parlent d'oscillations sinusoïdales)(Philipe, 2012). Ces oscillations issues de l'effet des vibrations rapides des corps, se propageant dans les milieux matériels et excitant l'organe de l'ouïe.

Quand un corps sonore a été frappé, ses différentes parties éprouvent aussitôt un mouvement de vibration. L'air qui environne ce corps participe à ce mouvement et forme autour de lui des ondes qui parviennent à l'oreille. L'air est donc le principal véhicule du son, qui se propage avec une vitesse de 340 m/s environ (à la température ordinaire)mais les sons aigus sont des vibrations plus rapides que les graves (leur fréquence est plus élevée).
Les liquides transmettent le son avec plus de rapidité ainsi que dans les solides, la vitesse est encore plus grande.

Le son ne se transmet pas dans le vide, et son intensité augmente ou diminue en même temps que la pression du gaz qui le transmet.
Lorsque les ondes sonores rencontrent des obstacles fixes, elles se réfléchissent. C'est sur cette propriété qu'est fondée la théorie de l'écho (Larousse, 1969).

Voici dans la figure 1.1comment vibre la membrane du tympan dans l'oreille en fonction du temps quand elle reçoit une note de musique(Philipe, 2012) :

-une note à 440 Hz qui est un son grave (un **La**, de la 3ème octave, en bleu).
- un son plus aigu à 1760 Hz(un **La** aussi, mais de la 5ème octave, en rouge).

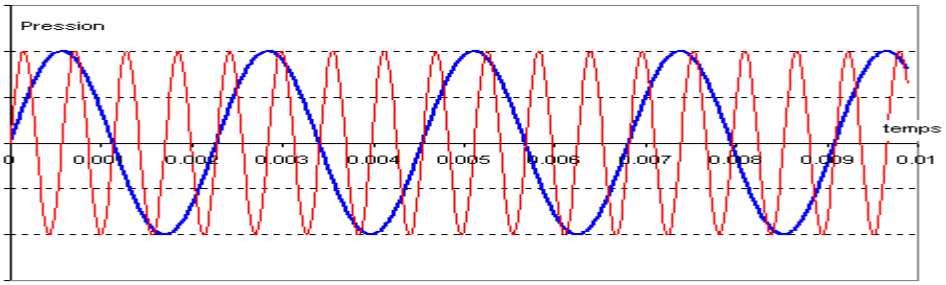


Figure1.1: Ondulation du tympan en fonction du temps pour sons grave et aigu(Philipe, 2012).

Ⅰ.3-Physiologie de l'oreille humaine

L'oreille est un organe jouant le rôle d'un transducteur mécanique quiconvertit l'information de pression en une information électrique(Khennouf, 2012). Elle est constituée chez l'être humain de 3 parties:

a. L'oreille externe, avec le pavillon et le conduit auditif, fermé par le tympan.
b. L'oreille moyenne, cavité ou caisse du tympan, communiquant avec le pharynx par la trompe d'Eustache, et dans laquelle une chaine de trois osselets (marteau, enclume, étrier) transmet les vibrations du tympan à la fenêtre ovale, et puis à l'oreille interne.
c. L'oreille interne, contient l'organe de l'équilibration et l'appareil auditif, formé de limaçon, contenant les cellules auditives ciliées de l'organe de Corti (Larousse,1969).

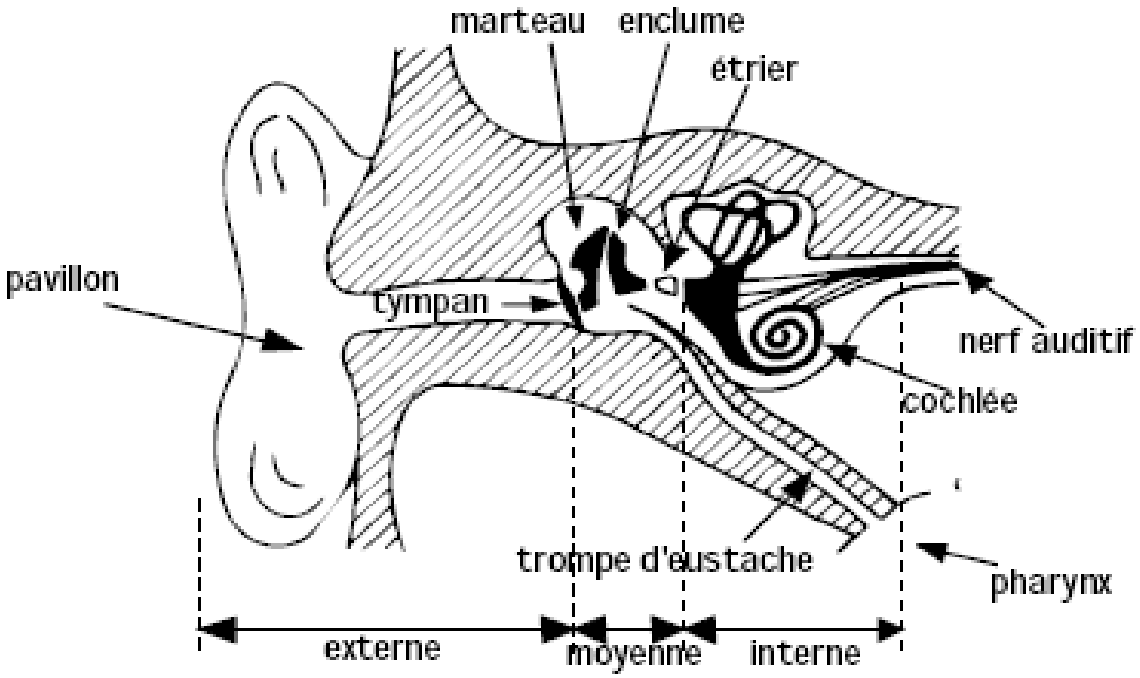


Figure 1.2: Système auditif humain (Ait Saadi , 2005).

Ⅰ.4-Mécanisme d'audition

Les ondes sonores sont recueillies par l'appareil auditif. Ces ondes de pression sont ensuite analysées dans l'oreille interne, puis les résultats sont interprétés par le cerveau.
Le conduit auditif relie le pavillon au tympan : c'est un tube acoustique de section uniforme fermé à une extrémité, son premier mode de résonance est situé à 3 KHz, ce qui accroît la sensibilité du système auditif humain dans cette gamme de fréquences.

Le mécanisme de l'oreille interne (*marteau,étrier, enclume*) permet une adaptation d'impédance entre l'air et le milieu liquide de l'oreille interne. Les vibrations de l'étrier sont transmises au liquide de la *cochlée*. Celle-ci contient la *membrane basilaire* qui transforme les vibrations mécaniques en impulsions nerveuses. La membrane s'élargit et s'épaissit au fur et à mesure que l'on se rapproche de l'apex de la cochlée; elle est le support de l'*organe de Corti* qui est constitué par environ 25000 *cellules ciliées* raccordées au nerf auditif. Chacune de ces cellules présente une réponse en fréquence où la fréquence de résonance dépend de la position occupée par la cellule sur la membrane.

L'oreille humaine est capable d'entendre sur une gamme de fréquence comprise entre les infrasons ($\leq$ 20 Hz) et les ultrasons ($\geq$ 20 kHz).À l'intérieur de cet espace fréquentiel existe un sous-espace délimité par les niveaux d'énergie des sons au-delà de lesquels l'homme ne percevra pas un son d'une fréquence appartenant au spectre de l'audition. Cette limite d'énergie est appelée seuil d'audition et il est variable en fonction de la fréquence. Inversement, il existe une limite d'énergie maximale appelée le seuil de douleur et qui varie aussi en fonction de la fréquence. Cette limite ne doit pas être franchie car la cochlée, et plus particulièrement les cellules ciliées, peuvent être irrémédiablement endommagées(Ait Saadi , 2005).

La présence des deux oreilles permet d'effectuer, au niveau du cerveau, des traitements plus complexes que le simple décodage d'une scène auditive. Le positionnement des oreilles de chaque côté du crâne permet en effet de profiter des capacités de la binauralité. Cette faculté permet de calculer la provenance d'un son en fonction du retard d'arrivée de ce dernier dans une oreille par rapport à l'autre. Il est à noter que cette binauralité permet à l'homme de discerner la position horizontale de l'émetteur d'un son mais pas sa position verticale(Benselama, 2007).

Ⅰ.5-La localisation de la source sonore

La perception des sons dans l'espace est essentiellement reliée à la perception binauriculaire, c'est-à-dire au fait d'avoir deux oreilles. La localisation des sons dans l'espace est complémentaire des autres sens, notamment la vision et la perception de son corps dans l'espace (proprio-perception).
La localisation comprend trois dimensions spatiales: l'azimut ou la direction (gauche-droite), la hauteur ou l'élévation (haut-bas), et la profondeur ou la distance à laquelle elle se trouve (Localisation d'un point d'émission,2012).

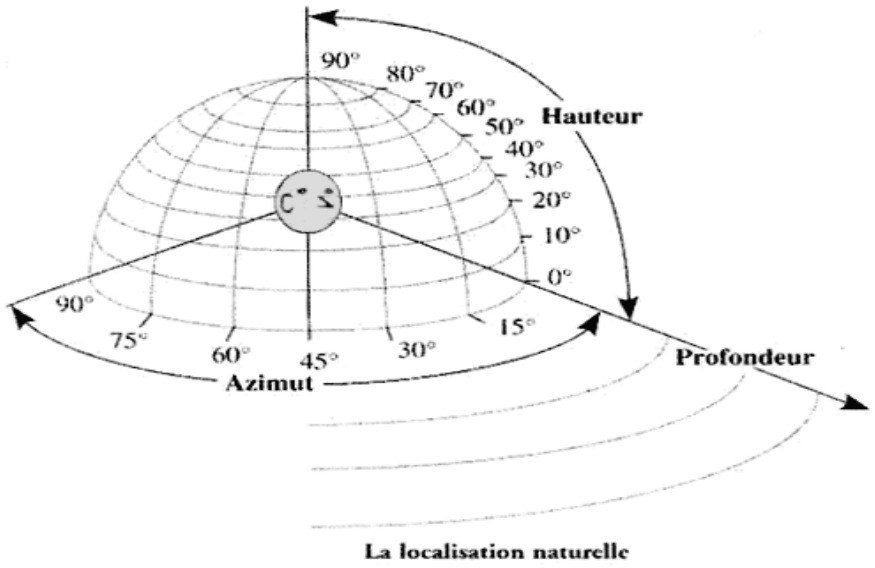


Figure1.3: La localisation naturelle de l'espace sonore (Localisation d'un point d'émission, 2012).

Ⅰ.5.1-Localisation à l'azimut (la direction)

Les indices pertinents sont les différences de niveau interauriculaires(DNI) et les différences de temps interauriculaires(DTI). Ces deux indices sont indépendants (localisation de la source sonore,2012).

a) Les Différences de Niveau Interauriculaires (DNI)

Il y a 120 ans, Lord Rayleigh avait déjà compris l'importance pour la localisation de la différence de niveaux d'intensité arrivant aux deux oreilles.
À cause du phénomène de la diffraction, un son venant de la droite sera plus fort à l'oreille droite qu'à l'oreille gauche, et ce d'autant plus que la fréquence est élevée. En effet, la tête constitue un obstacle à franchir pour les ondes sonores et une "ombre acoustique" apparaît, d'autant plus marquée que la fréquence est élevée(localisation de la source sonore, 2012).

- Calcul de la fréquence critique (correspondant aux dimensions d'une tête) :

$$\text{Diamètre de la tête} = 23\ \text{cm} = 0{,}23 \qquad (1.1)$$

$$f = c / l = 344 / 0{,}23 \approx 1500\ \text{Hz} \qquad (1.2)$$

À 500 Hz, la longueur d'onde d'un son est de 68 cm, ce qui représente plus de 3 à 4 fois le diamètre d'une tête moyenne. À cause de la diffraction, la DNI est donc petite pour des fréquences inférieures à 500 Hz, si la source est au moins à un mètre de distance.

- La DNI est un indice de localisation surtout utilisé pour les hautes fréquences (au-delà de 4000 Hz).

La figure1.4 représente la DNI en fonction de l'angle d'azimut mesuré pour différentes fréquences. On remarque que la différence de niveau s'accentue pour des plus hautes fréquences.

(Localisation de la source sonore, 2012).

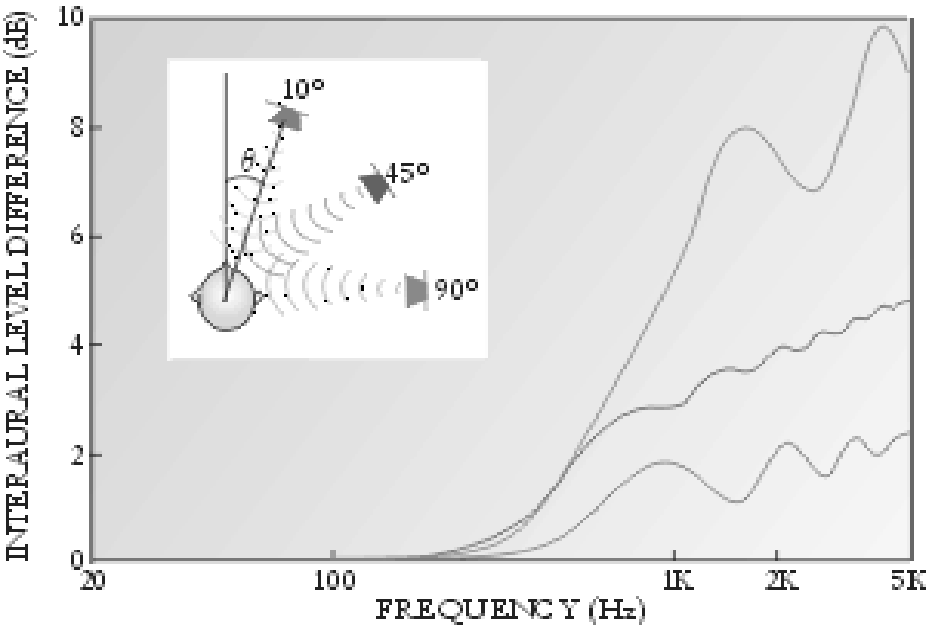


Figure 1.4: La DNI en fonction de l'angle d'azimut pour différentes fréquences(localisation de la source sonore, 2012).

a) Les Différences de Temps Interauriculaires (DTI)

Pour les basses fréquences (grandes longueurs d'ondes →λ > tête), il semble que notre système auditif binauriculaire utilise la différence de phase (Δ φ) ou différence de temps d'arrivée (Δ t) entre les signaux arrivant aux deux oreilles (les signaux étant repérés par les maxima ou les passages à zéro) (localisation de la source sonore, 2012,).

- Une différence de phase Δ φ correspond à une DTI :

$$\Delta t = \Delta\varphi/(2\pi f) \qquad (1.3)$$

- Pour calculer une approximation de la DTI, on considère la demi-circonférence correspondant au parcours de l'onde autour de la tête pour passer d'une oreille à l'autre :

$$d = \pi R \qquad (1.4)$$

OùR est le rayon de la tête.

- La DTI vaut alors :

$$\Delta t = d/V \qquad (1.5)$$

Où v est la vitesse du son dans l'air.

- Calcul de la DTI :

Rayon de la tête = 10 cm = 0,1 m (1.6)

Demi-circonférence = 0,1 π = 0,314 m (1.7)

La plus grande valeur possible (dans le cas où la source est en face d'une des deux oreilles) est :

DTI = Δ t = 0,314 / 340= 924 μs = 0,924 ms (1.8)

Pour les grandes longueurs d'onde, on peut utiliser la formule pour la diffraction par une sphère :

$$\Delta t = \left(\frac{3R}{V}\right) \sin\theta \qquad (1.9)$$

Où θ est l'angle azimutal (gauche-droite).

À 90° (source en face d'une des deux oreilles) :

Δ t = (3 x 0,1 / 340) = 0,882 ms. (1.10)

- Les DTI sont des indices de localisation pour les plus basses fréquences(≤1500 Hz).

(Localisation de la source sonore,2012).

Ⅰ.5.2-Localisation de la profondeur (l'élévation)

Considérant le modèle de la tête sphérique, tous les points compris dans un plan vertical perpendiculaire à l'axe passant par les deux oreilles, correspondent à des signaux arrivant aux deux oreilles avec la même intensité et la même phase. Les indices de DNI et DTI ne permettent donc pas de distinguer le haut, le bas, l'avant et l'arrière, dans le plan vertical.
D'autre part, des sources sonores situées en tout point de la surface d'un cône, appelé cône de confusion(Figure 1.5), causeront des DTI identiques. Sans le support visuel, le son peut être perçu comme provenant du côté opposé sur le cône, par rapport à son point d'origine réel(localisation de la source sonore,2012).

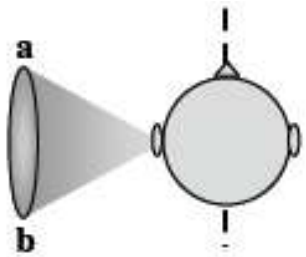


Figure 1.5: Le cône de confusion (localisation de la source sonore,2012).

Malgré tout, la localisation avant/arrière et haut/bas est possible grâce à l'interprétation par le cerveau du filtrage acoustique du son par la Fonction de Transfert Anatomique (FTA)ou en anglais Head-Related Transfer Function (HRTF), correspondant au filtrage par le pavillon, la tête et les épaules. La FTA est illustrée dans la figure 1.6 pour trois positions dans le plan vertical (localisation de la source sonore,2012).

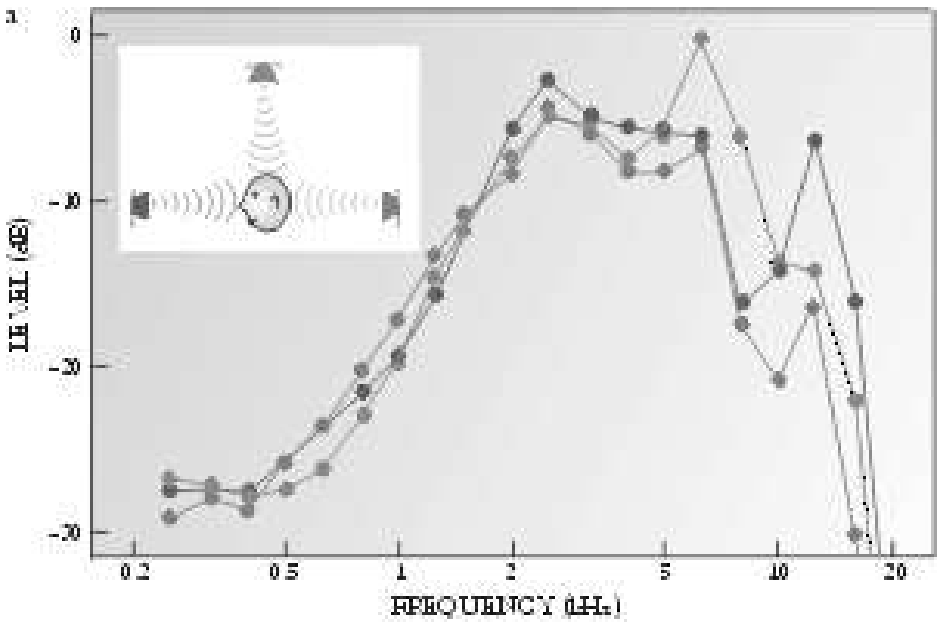


Figure 1.6: La Fonction de Transfert Anatomique (localisation de la source sonore, 2012).

Ⅰ.5.3-La perception de la distance de la source

Lorsqu'on s'éloigne d'une source sonore, l'intensité diminue de 6 dB à chaque doublement de distance.
Diverses expériences ont démontré que des sources sonores familières sont localisées avec plus de précision que des sources inconnues mais ces distances sont toujours sous-estimées. L'estimation de la distance dépend également du niveau d'émission de la source.
En conséquence, si l'on désire recréer artificiellement une impression de profondeur, une atténuation supérieure à 6 dB (jusqu'à 20 dB) est nécessaire pour donner l'illusion d'un doublement de distance(localisation de la source sonore,2012).

a) **En champ libre**

L'analyse spectrale par le système auditif est le principal indice pour la perception de la distance (si la référence connue). En effet, la répartition des graves et des aigus changes avec la distance. Cette faculté est plus précise si la source est en mouvement (localisation de la source sonore, 2012).

b) **Dans un espace réverbérant**

Dans une salle, lorsqu'un auditeur s'éloigne d'une source sonore, le rapport du son direct au son réverbéré diminue.
Le son réverbéré est constitué de multiples réflexions ; il devient rapidement homogène en tous points de la salle. En fait, seule l'énergie du son direct décroît de 6 dB à chaque doublement de distance, d'où une même décroissance du rapport champ direct / champ réverbéré(Localisation d'un point d'émission, 2012).

Notre perception de la distance peut aussi être influencée par notre expérience avec un type de source particulier tel que la voix. Dans une expérience, on avait présenté des voix criées, parlées, et chuchotées à des pressions sonores identiques. Les voix criées étaient systématiquement évaluées comme étant plus distantes, suivies des voix parlées, puis des voix chuchotées(localisation de la source sonore,2012).

La densité spectrale d'un signal sonore varie lors de sa propagation en fonction d'une absorption inégale des fréquences graves et aiguës.
Des expériences ont montré notamment qu'en champ libre, des sons dont le contenu fréquentiel est inférieur à 2 kHz semblent plus éloignés que les sons de fréquence supérieure
La figure 1.7ci-dessous présente les résultats de l'estimation de la distance d'une voix parlée masculine chuchotée, à spectre essentiellement aigu, comparée à celle d'une voix parlée normalement à large spectre. On constate que la perception de la distance d'une voix chuchotée est sous-estimée.
On en déduit que la sensation d'une source sonore proche ou lointaine pourra être créée artificiellement par un filtrage approprié(Localisation d'un point d'émission,2012).

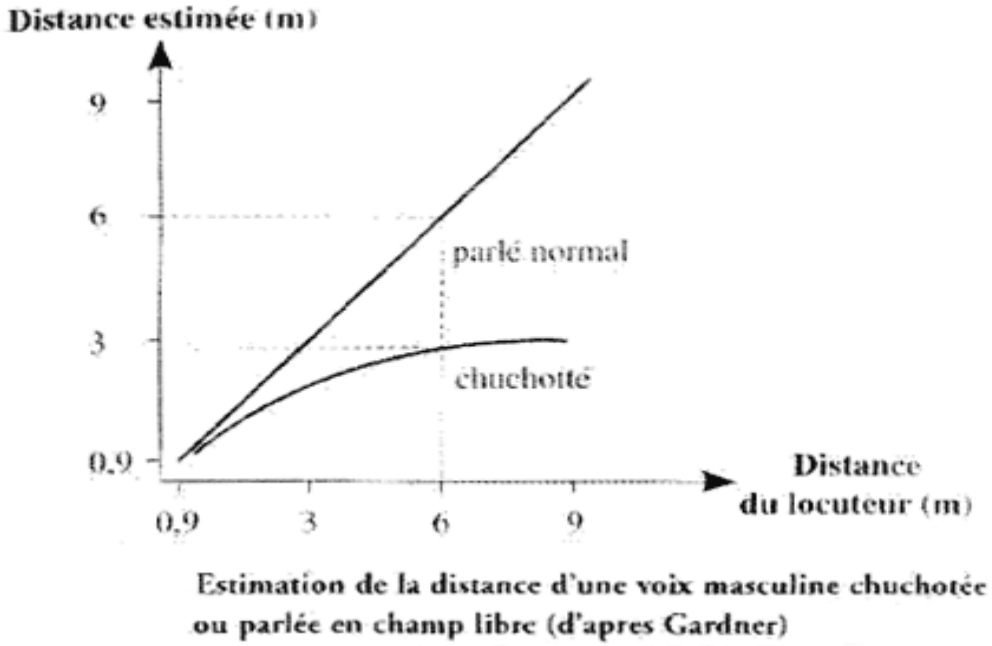


Figure 1.7: Estimation de la distance d'une voix parlée ou chuchotée en espace libre

Ⅰ.6- Conclusion

Dans ce premier chapitre, on a juste jeté un coup d'œil sur le système auditif humain et le mécanisme d'audition ainsi on a rappelé de quelques théories sur lesquelles est basée la localisation de sources sonores.

II.1- Introduction

Dans ce chapitre, on va proposer quelques méthodes de localisation de sources sonores telles que : la corrélation et en particulier la corrélation filtrée, la méthode du différentiel d'énergie et enfin la GCC-PHAT.

L'objectif de toutes ces méthodes est d'aboutir aux mêmes performances que le système auditif humain en termes d'extraction, reconnaissance et localisation.
La comparaison de ces différentes techniques de localisation proposées, nous permet d'établir les avantages et les inconvénients d'une approche donnée dans une mise en œuvre en temps réel **(**Brutti, 2008**).**

II.2-Techniques de localisation de la position

La localisation peut se faire soit par le calcul de la différence d'énergie des signaux arrivant aux deux microphones, ou par le calcul du retard entre les deux signaux de parole collectés par les deux microphones puis déduire la distance entre le locuteur et les deux microphones.

II.2.1- Méthode du différentiel d'énergie (EDM)

a) Calcul du différentiel d'énergie

- Soit un signal s(n) définie sur un intervalle de taille **N** c-a-d [1 N], l'énergie du signal s(n) dans le cas discret peut s'écrire comme suit :

$$E_s=\sum_1^N |S_i|^2 \qquad (2.1)$$

- Dans notre application, on a utilisé l'énergie modifiée qui est calculée comme étant la somme des valeurs absolues des valeurs du signal le long d'une fenêtre d'analyse de taille **N**. Cette énergie est appelée : Energie d'ordre 1 (Khennouf, 2010).

- Les deux énergies (énergie du signal perçu par le microphone droit et l'énergie du signal perçu par le microphone gauche) sont données par les relations suivantes:

$$E_{Sd}=\sum_1^N |X_i| \qquad (2.2.a)$$

$$E_{Sg}=\sum_1^N |Y_i| \qquad (2.2.b)$$

- L'énergie des deux signaux en dB est donnée par :

$$E_{Sd}(dB) = 10*\log_{10}(E_{Sd}) \quad (2.3.a)$$

$$E_{Sg}(dB) = 10*\log_{10}(E_{Sg}) \quad (2.3.b)$$

- Le différentiel de l'énergie logarithmique DE est :

$$DE = E_{Sd}(dB) - E_{Sg}(dB) \quad (2.4)$$

b) Indication de la position localisée

Pour la méthode **EDM,** le principe est simple afin d'estimer la position relative du locuteur par rapport à la position des microphones.

Après avoir calculé la différence d'énergie logarithmique DE, on compare cette différence à un seuil de détection pour localiser le locuteur qui parle.

Ceci se fait suivant l'algorithme ci-dessous:

```
% indication sur la position selon la valeur de DE

Si DE > seuil_max
pos = 'droite'

Sinon Si DE < seuil_min
pos = 'gauche'

Sinon
pos = 'milieu'
Fin Si;
```

Avec les valeurs de **$seuil_{max}$** et de **$seuil_{min}$** sont déterminées expérimentalement.

II.2.2- Méthodes basées sur le calcul du retard

Comme deuxième technique de localisation, on propose le calcul de la distance entre la source sonore (le locuteur qui parle dans notre étude) et les deux microphones placés l'un à droite du locuteur et l'autre à gauche.
Le calcul de cette distance est basé essentiellement sur le calcul du retard.

II.2.2.1. Calcul de la distance entre le locuteur et les deux microphones

Pour cela, on considère un locuteur entre deux microphones espacés tel qu'il est illustré dans la figure 2.1, dont les différents paramètres 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.
- $\boldsymbol{L}$ **:** La distance entre les deux microphones.
- $\boldsymbol{V_S}$**:** La vitesse du son qui est environ 340 m/s dans l'air et à une température ordinaire de 15°C

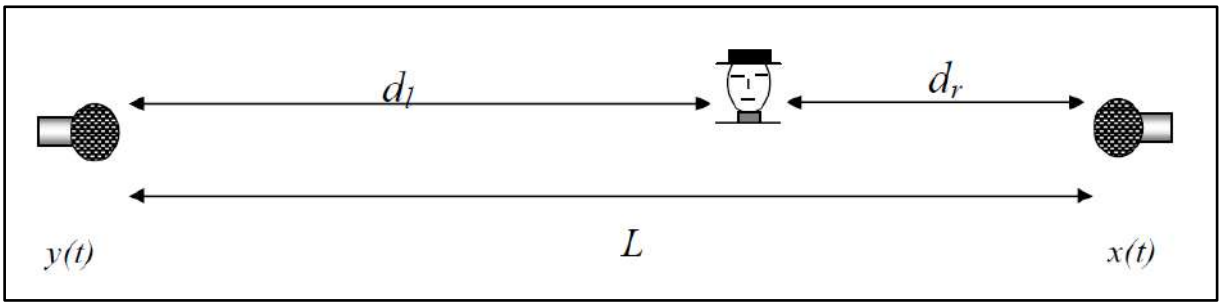


Figure 2.1: Schéma représentant un locuteur entre deux microphones (Khennouf, 2010).

Par exemple, on calcule la distance d_r entre le locuteur et le microphone droit :

$$\mathrm{d_r} + \mathrm{d_1} = \mathrm{L} \quad (2.5)$$

$$\mathrm{delay} = (\mathrm{d_r} - \mathrm{d_1})/\mathrm{V_S} \quad (2.6)$$

$$\mathrm{d_r} - \mathrm{d_1} = \mathrm{delay} * \mathrm{V_S} \quad (2.7)$$

$$\mathrm{d_r} = (\mathrm{L} + \mathrm{delay} * \mathrm{V_S})/2 \quad (2.8)$$

Selon cette distance calculée, on peut détecter la position du locuteur qui parle.

Ceci se fait suivant l'algorithme ci-dessous:

```
%indication sur la position

Si d<seuil_min
pos='droite'

Sinon Si d>seuil_max
pos='gauche'

Sinon
pos='milieu'
Fin Si;
```

Avec les valeurs de **$seuil_{max}$** et de **$seuil_{min}$** sont déterminées expérimentalement.

II.2.2.2.Calcul du retard

a) Méthode de la corrélation(CM)

a.1) Définition

La corrélation est un concept issu de la biologie. C'est par le biais des travaux de Francis Galton (Corrélation, wikipédéa, 2012) que la corrélation devient un concept statistique.

En probabilités et en statistiques, étudier la **corrélation** entre deux ou plusieurs variables aléatoires c'est étudié l'intensité de la liaison qui peut exister entre ces variables (Corrélation, wikipédéa, 2012)

La fonction de corrélation temporelle est définie par la mesure de similitude entre deux signaux, dont l'un est décalé dans le temps par rapport à l'autre. On distingue deux types de corrélation : l'autocorrélation et l'intercorrélation (Khennouf, 2010).

a.2) L'autocorrélation

L'autocorrélation est un outil mathématique souvent utilisé en traitement du signal, consiste à comparer un signal par lui-même décalé avec un certain retard **τ**. L'autocorrélation

permet de détecter des régularités, des profils répétés dans un signal comme un signal périodique perturbé par beaucoup de bruit (autocorrelation, wikipédéa, 2012).

La fonction d'autocorrélation est définie par la relation suivante :

$$cor_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t+\tau)dt \qquad (2.9)$$

a.3) L'inter-corrélation « cross corrélation »

De sa part, l'inter-corrélation est un outil mathématique qui consiste à comparer deux signaux différents dont l'un est décalé avec un certain retard $\boldsymbol{\tau}$ par rapport à l'autre.

La fonction d'intercorrélation est définie par la relation suivante :

$$cor_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)y(t+\tau)dt \qquad (2.10)$$

a.4) Estimation du retard par l'inter-corrélation « cross corrélation »

L'inter-corrélation était la première approche pour estimer le retard entre les signaux reçus par les deux microphones.
Dans le cas discret, la fonction de cross corrélation est donnée par la relation suivante :

$$cor_{dg}(\tau) = \sum_{1}^{N} S_d(n) S_g(n+k) \qquad (2.11)$$

Avec : $S_d(n)$: Signal enregistré par le microphone droit.

$S_g(n)$: Signal enregistré par le microphone gauche.

$$k = \tau . F_e \qquad (2.12)$$

Où F_e : est la fréquence d'échantillonnage.

La figure 2.2 montre la structure de l'inter-corrélation.

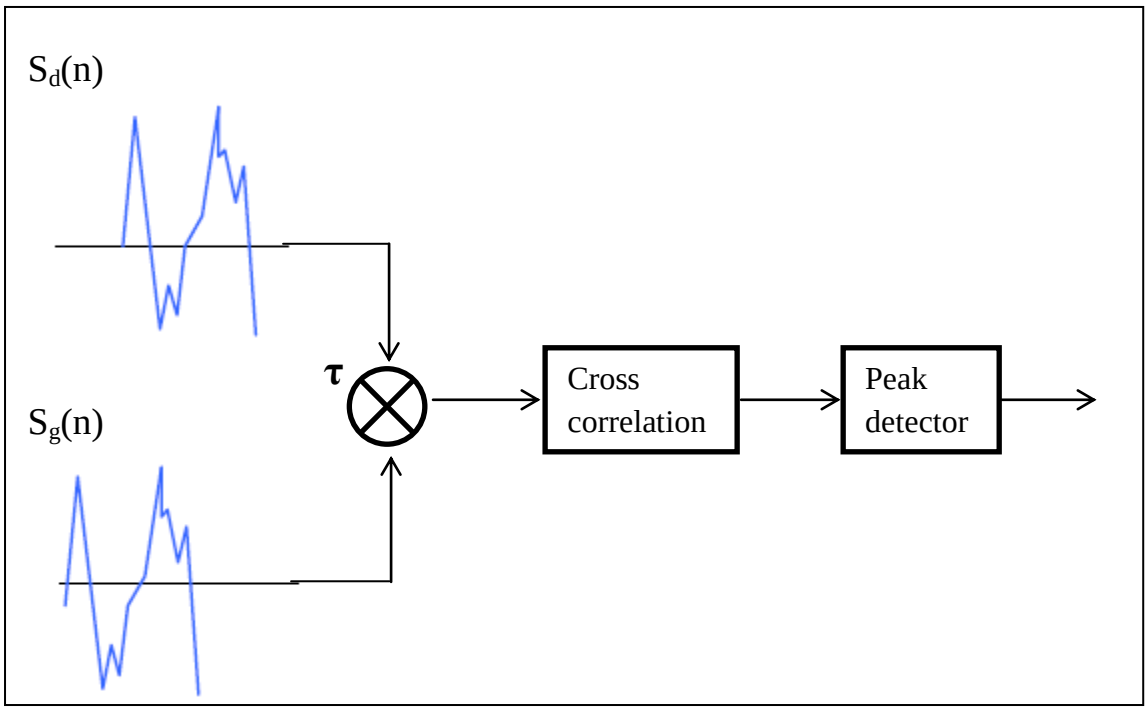


Figure 2.2 : Estimation du retard par la méthode de corrélation.

- L'emplacement (indice) du pic maximal dans la sortie représente le retard estimé :

$$\tau = argmax(cor_{dg}(k)) \quad (2.13)$$

b) Méthode de la corrélation filtrée (FCM)

La méthode de corrélation filtrée **FCM** se fait par deux étapes :

- On calcul d'abord la corrélation discrètetelle quelle est calculée par la relation (2.11).
- Ensuite, on applique au signal $cor_{dg}(\tau)$un filtre passe-bande elliptique, dont la syntaxe est comme suit:

$$f = ellip(N, R_p, R_s, F_p) \quad (2.14)$$

Avec :

N : ordre du filtre.

Rp : atténuation maximale (en dB) dans la bande passante.

Rs : atténuation minimale (en dB) dans la bande coupée.

Fp =2*fc où fc : est la fréquence de coupure.

La figure 2.3 montre la structure de la corrélation filtrée FCM :

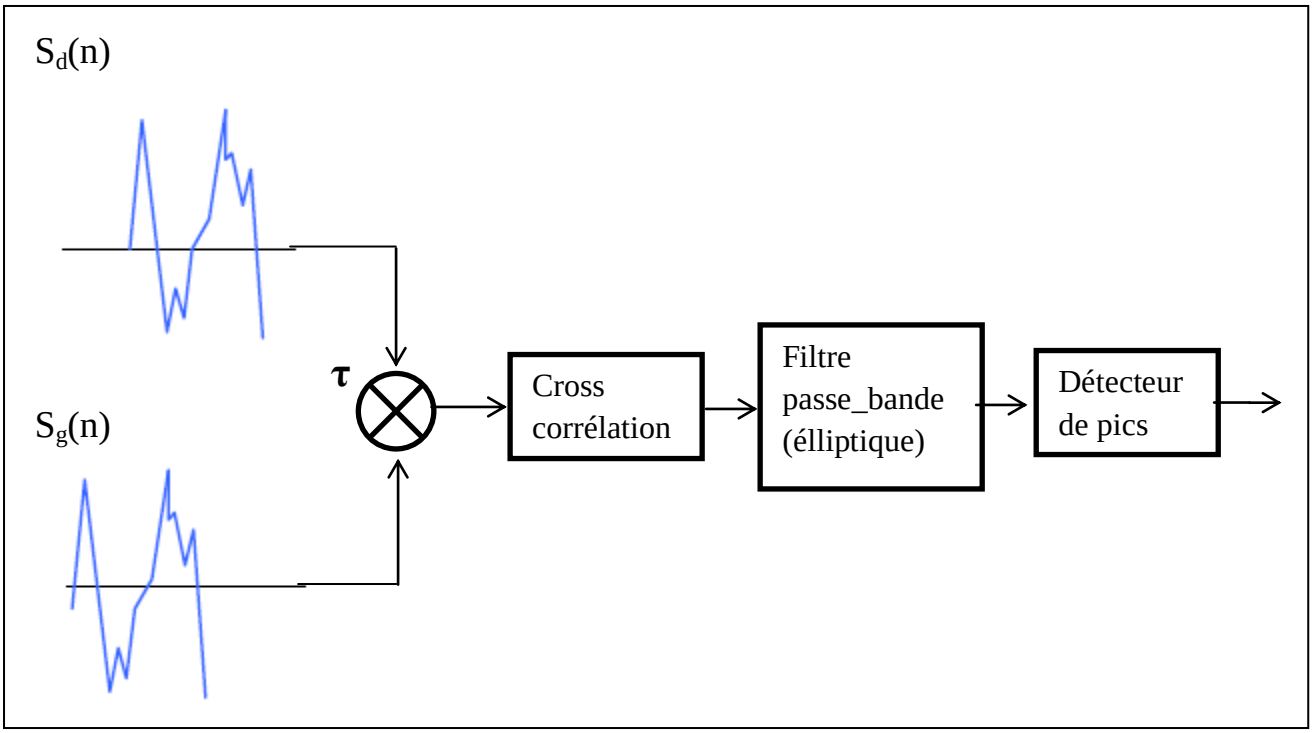


Figure 2.3 : Estimation du retard par la méthode de corrélation filtrée.

- L'emplacement (indice) du pic maximal dans la sortie représente le retard estimé :

$$\tau = argmax(fcor_{dg}(k)) \quad (2.15)$$

c) Méthode de Generalized Cross-Correlation PHase Transform (GCC –PHAT)

c.1) Définition

La GCC-PHAT est parmi les méthodes les plus utilisées pour déterminer la différence de temps d'arrivée (TDOA) «Time Difference Of Arrival» entre deux microphones. Donc, à l'aide de plusieurs valeurs de TDOA, on peut estimer l'emplacement de la source.

Par exemple : on considère un système composé d'une source et 4 microphones.

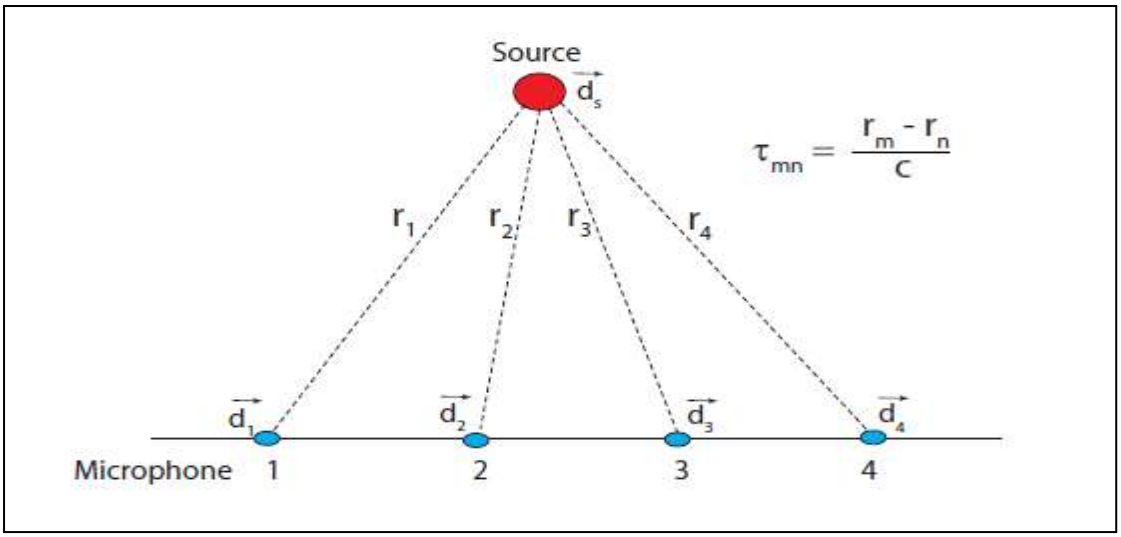


Figure 2.4: TDOA entre chaque deux microphone.

Si la distance entre la source et le microphone est r_m(m=1, 2, 3 ,4) ; le temps de parcours du signal sonore de la source au microphone est :

$$\tau_m = \frac{r_m}{V_S} \qquad (2.16)$$

Où$V_S$est la vitesse du son qui est environ 340 m/s dans l'air et à une température ordinaire de 15°C

Donc, le temps de retard TDOA entre deux microphones m et n peut être défini comme :

$$\tau_{mn} = \tau_m - \tau_n = \frac{r_m - r_n}{V_S} \qquad (2.17)$$

De cette relation entre la TDOA et les distances entre la source et les microphones ,r_m, on peut estimer l'emplacement source**(**Hoang, 2009).

c.2) Dérivation de la GCC

Considérons deux signaux $S_d(t)$ et $S_g(t)$ (signaux reçus respectivement par le microphone droit et le microphone gauche), dont $S_g(t)$ est retardé d'une valeur $\boldsymbol{\tau}$ par rapport à $S_d(t)$.
Donc, on peut écrire :

$$S_g(t) = S_d(t + \tau) \qquad (2.18)$$

Calculons maintenant la différence du temps d'arrivée$\boldsymbol{\tau}$, entre les deux microphones.

- L'inter-corrélation des deux signaux sera montrée un pic correspondant à$\boldsymbol{\tau}$.

Cette inter-corrélation est définie comme suit :

$$cor_{dg}(\tau) = \int_{-\infty}^{+\infty} S_d(t)\ S_g(t)dt = \int_{-\infty}^{+\infty} S_d(t)\ S_d(t+\tau)dt \qquad (2.19)$$

- On prend la transformée de Fourier de l'inter-corrélation $COR_{dg}(\omega)$

$$COR_{dg}(\omega) = \int_{-\infty}^{+\infty} cor_{dg}(t)e^{j\omega\tau}\ \ d\tau \qquad (2.20)$$

- Appliquons la propriété de la convolution de transformée de Fourier sur (2.18) lors de sa substitution en (2.19), on aura :

$$COR_{dg}(\omega) = S_d(\omega).S_g^*(\omega) \qquad (2.21)$$

Avec :

$S_d(\omega)$et$S_g(\omega)$: sont respectivement les transformée de Fourier des signaux $S_d(t)$et $S_g(t)$. $S_g^*(\omega)$est le conjugué de $S_g(\omega)$.

- La transformée de Fourier inverse de(2.20) nous donne la fonction de l'inter-corrélation

$$cor_{dg}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_d(\omega).S_g^*(\omega)e^{j\omega\tau}\ \ d\omega \qquad (2.22)$$

- L'inter-corrélation généralisée GCC de $S_d(t)$et $S_g(t)$ est l'inter-corrélation de leurs deux versions filtrées.
  En désignant les transformées de Fourier de ces deux filtres $F_d(\omega)$ et$F_g(\omega)$, on aura la GCC,$R_{dg}(\tau)$qui est défini comme,

$$R_{dg}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}(F_d(\omega).S_d(\omega)).(F_g(\omega).S_g(\omega))^* e^{j\omega\tau}\ \ d\omega \qquad (2.23)$$

On définit une fonction de pondération combinée $\Psi_{dg}(\omega)$comme,

$$\Psi_{dg}(\omega) = F_d(\omega).F_g^*(\omega) \qquad (2.24)$$

En remplaçant (2.23) dans (2.24), la GCC devient :

$$R_{dg}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Psi_{dg}(\omega).S_d(\omega).S_g^*(\omega)e^{j\omega\tau}\ d\omega \qquad (2.25)$$

(Hoang, 2009)

c.3) Phase Transform (PHAT)

Le TDOA entre les deux microphones est le laps de temps $\boldsymbol{\tau}$ qui maximise la GCC, $\boldsymbol{R_{dg}(\tau)}$.
En réalité,$\boldsymbol{R_{dg}(\tau)}$a de nombreux maxima locaux rendant ainsi plus difficile à détecter la vraie valeurs maximale.
Le choix des fonctions de pondération, $\Psi_{dg}(\omega)$agit sur la performance de la GCC.

Il a été montré que la fonction de pondération de phase transform(PHAT) étudiée dans des conditions réverbérant-libre est robuste dans des environnements réalistes. Cette fonction PHAT qui donne le maximum optimal est définie comme suit,

$$\Psi_{dg}(\omega) = \frac{1}{|S_d(\omega).S_g^*(\omega)|} \qquad (2.26)$$

(Hoang, 2009)

c.4) La GCC - PHAT

En remplaçant la fonction de pondération PHAT de l'équation (2.25) dans l'expression de la GCC de l'équation(2.24), on obtient la GCC-PHAT qui est définie par :

$$R_{dg}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{1}{|S_d(\omega).S_g^*(\omega)|} S_d(\omega).S_g^*(\omega)e^{j\omega\tau}\ d\omega \qquad (2.27)$$

Ou bien, en d'autre terme: [brut]

$$R_{dg}(\tau) = ifft\left(\frac{S_d(\omega).S_g^*(\omega)}{|S_d(\omega)||S_g(\omega)|}\right) \quad (2.28)$$

(Hoang, 2009)

La figure 2.5 montre la structure de la GCC-PHAT :

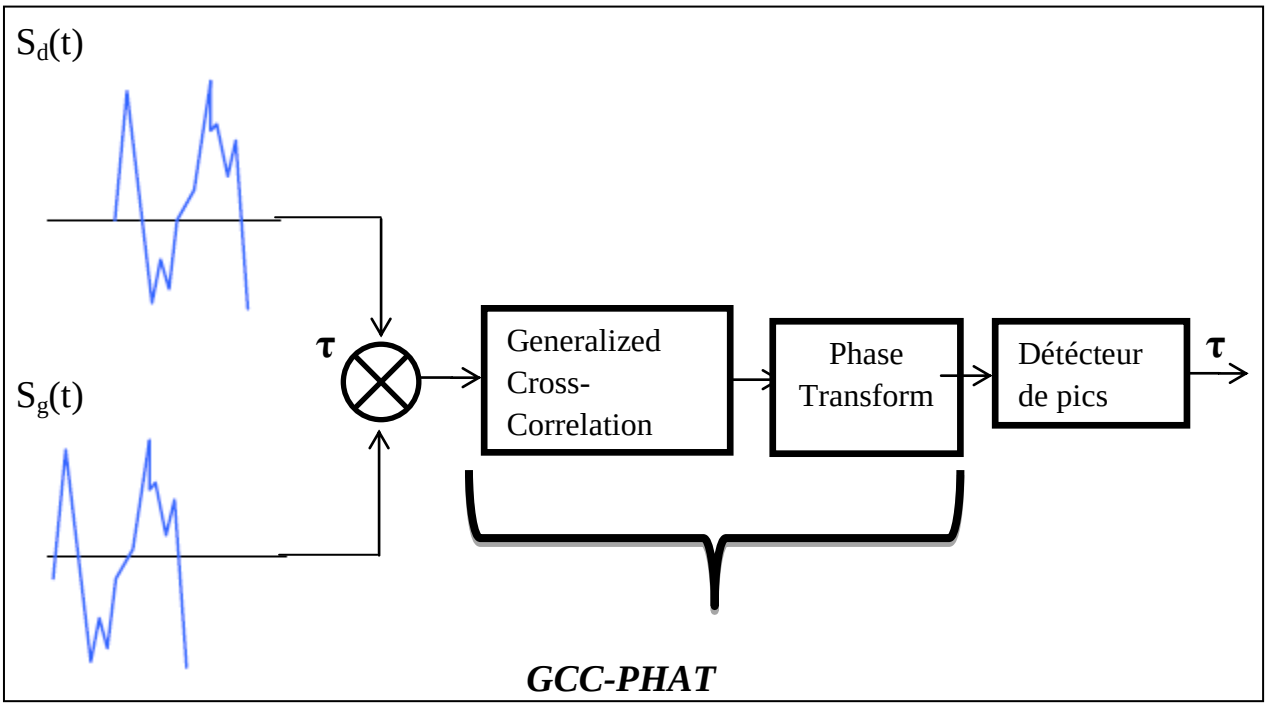


Figure 2.5 : Estimation du retard par la méthode de GCC-PHAT.

- L'emplacement (indice) du pic maximal dans la sortie représente le retard estimé

$$\tau = argmax(R_{dg}(\tau)) \quad (2.29)$$

II.3- Exemple d'application de localisation

- **Cas d'un signal sinusoïdal (cas théorique)**

Avant d'appliquer les différentes méthodes sur notre signal de parole (Voir chapitre 4) et afin d'étudier leurs performances, on a proposé de les tester d'abord sur deux sinusoïdes légèrement décalées dans le temps.

a) Méthode du différentiel d'énergie EDM

On génère de la même source deux sinusoïdes de même période et de même amplitude.

Exemple :

Le capteur de droite reçoit le signal :

a=sin(2*pi*t*fo).*exp(-0.0005*t)+bruit;

Le capteur du gauche reçoit le même signal mais qui a subi une atténuation

b=0.4*(sin(2*pi*t*fo).*exp(-0.0005*t)+bruit);

Leurs énergies seront calculées comme suit :

$E1=\sum_1^N |a(t)|^2$

$E2=\sum_1^N |b(t)|^2$

- L'énergie des deux signaux en dB est donnée par :

$$E1(dB) = 10*\log_{10}(E1)$$

$$E2(dB) = 10*\log_{10}(E2)$$

- Le différentiel de l'énergie logarithmique DE est :

$$DE = E1(dB) - E2(dB)$$

Selon le signe de DE qu'on détermine la position de la source.

- Après simulation, on a trouvé : **DE= +8,399> 0(** E1(dB) >E2(dB)**)**,

Donc la source est plus proche du capteur qui se trouve au coté droit.

b) Méthodes basées sur le calcul du retard (CM, FCM, GCC-PHAT)

On génère de la même source deux sinusoïdes de même période et de même amplitude mais légèrement décalées dans le temps (décalage =23 échantillons).

```
a=sin(2*pi*t*fo).*exp(-0.0005*t)+bruit;
b=sin(2*pi*(t-decalage)*fo).*exp(-0.0005*(t-decalage))+bruit_decale;
```

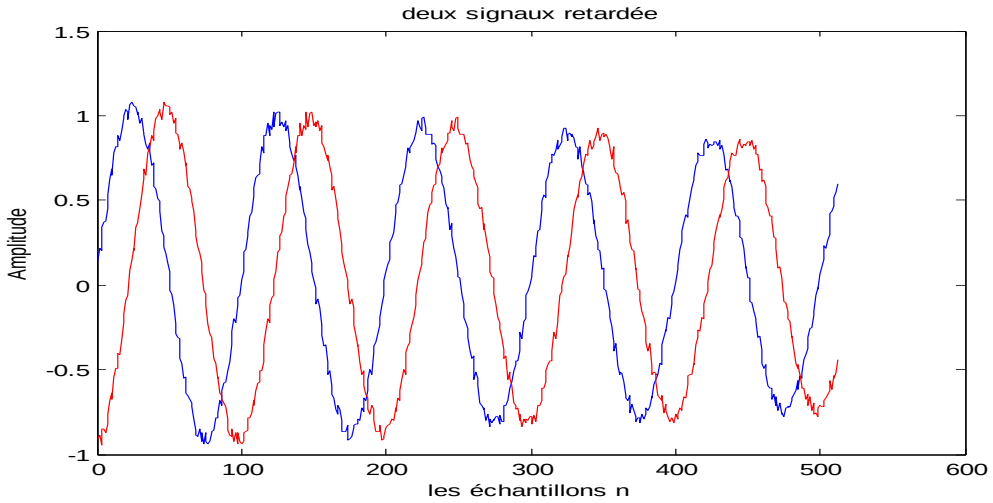


Figure 2.6 : Deux sinusoïdes légèrement décalées.

b.1) Calcul du retard

- Pour la méthode de la corrélation **CM** : on a utilisé l'instruction MatLab c=xcorr(a,b)
- Pour la méthode de la corrélation filtrée **FCM** : on a filtré le résultat de la méthode précédente par un filtre passe-bande de type elliptique [f=ellip (N, Rp, Rs, Fp)]

Avec :

N : ordre du filtre. (On prend N=4)

Rp : atténuation maximale (en dB) dans la bande passante. (On prend Rp=0.5 dB)

Rs : atténuation minimale (en dB) dans la bande coupée. (On prend Rs=60 dB)

Fp =2*fc avec fc=0.25Hz (0<fc<fe/2 et fe=1Hz)

Donc :

f=ellip(4,0.5,60,0.5);

- Pour la méthode de la **GCC-PHAT** : on a utilisé la formule(2.28).

La figure 2.7 illustrent les différentes méthodes de localisation :

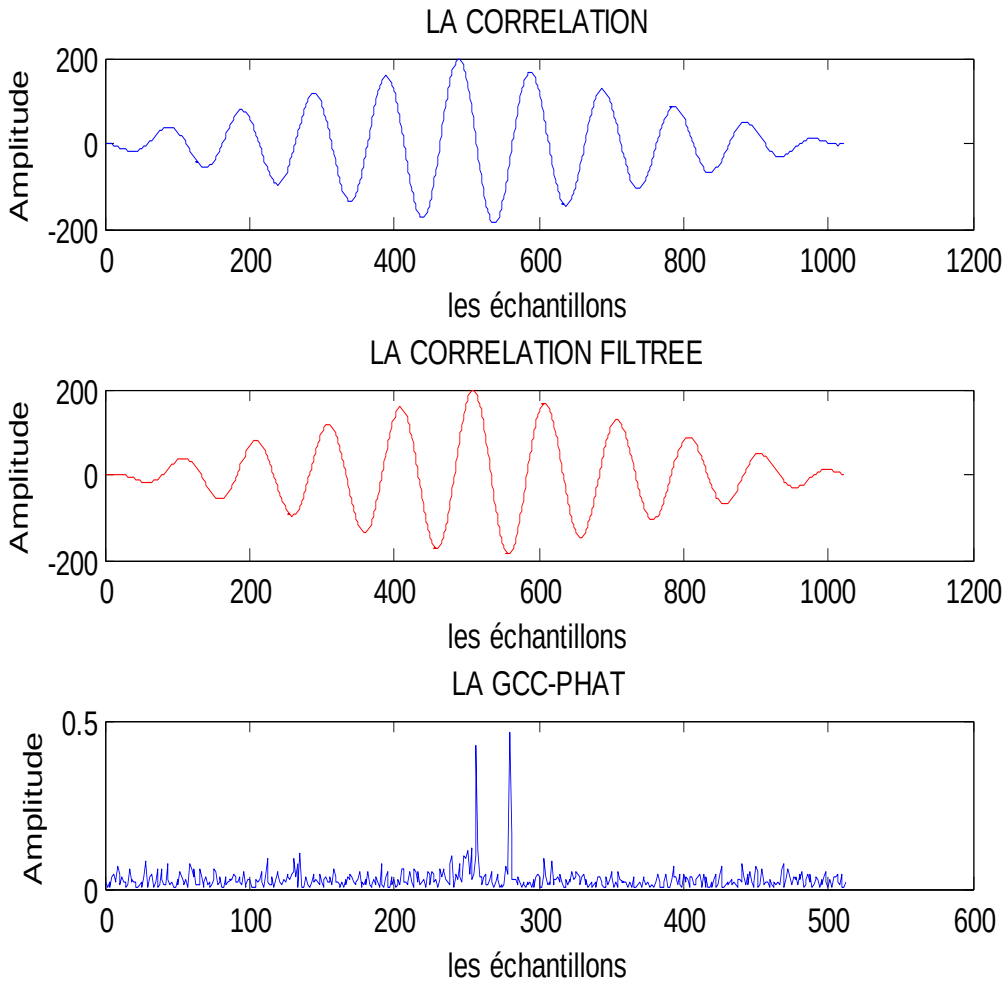


Fig 2.7: L'allure des différentes techniques pour le calcul du retard.

- L'emplacement du pic maximal dans les différents graphes représente le retard estimé.

Après simulation par MatLab, on a obtenues les résultats illustrés dans le tableau suivant :

Tableau 2.1: Résultats de simulation des différentes méthodes.

La méthode	**Le retardτ (échantillon)**
CM	23
FCM	23
GCC-PHAT	23

Commentaire

Théoriquement, toutes les méthodes présentent une bonne performance dans le calcul du retard entre les deux signaux.

b.2) Localisation de la position

Comme on a démontré dans le paragraphe (**II.2.2.1.)** de ce chapitre, calculer la distance revient à calculer le retard. Cette distance va nous indiquer sur la position de la source selon la relation (2.8).

Avec :

$$\tau \text{ [Second]} = \tau \text{ [échantillon]} / Fe \qquad (2.30)$$

Où Fe est la fréquence d'échantillonnage.

II.4- Conclusion

Dans ce chapitre, on a proposé quelques méthodes utilisées pour la localisation des locuteurs telles que la corrélation filtrée, le différentiel d'énergie et la GCC-PHAT.
Afin d'étudier leurs performances, on les a appliqué d'abord sur un signal sinusoïdal et après avoir les testé, on a trouvé que pratiquement toutes les techniques ont donné des très bons résultats de localisation.

III.1- Introduction

Pour les tests d'évaluation des performances de notre système, la conception d'une base de données constituée de plusieurs scénarios avec différentes configurations était nécessaire. Les enregistrements audio ont été faits à l'aide d'un systèmed'acquisition réalisé.
Dans ce chapitre, on va présenter toutes les étapes, le matériel et les logiciels utilisée pour réaliser et traiter cette base de données.

III.1- Expériences réalisées

L'objectif de notre travail est la conceptiond'un système automatique intelligent afin d'orienter une caméra mobile vers des cibles sonores (des locuteurs qui parlent dans notre application).
Ces locuteurs sont assis derrière une table comme illustré dans la figure 3.1suivante :

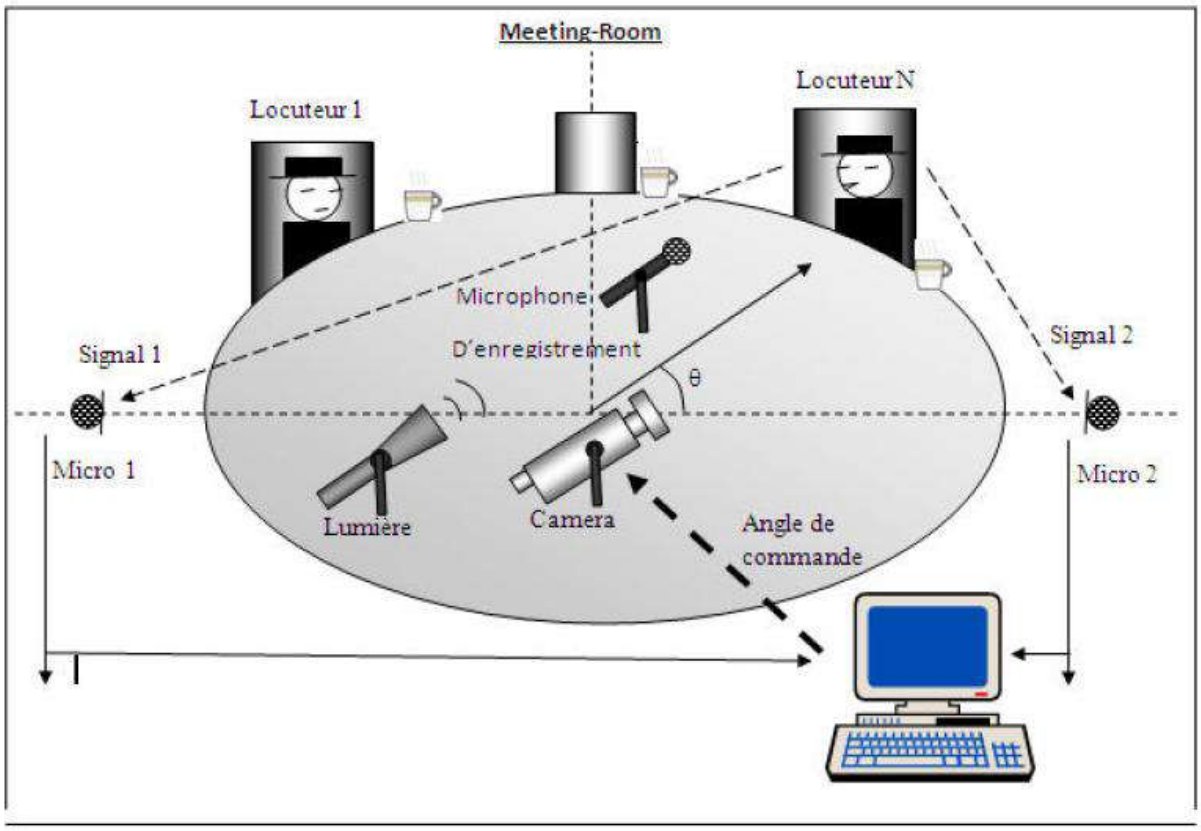


Figure 3.1:Schéma synoptique de la configuration de la meeting-room.

Plusieurs expériencesont été effectuéesenchangeant à chaque fois différents paramètres dans le but de réaliser une base de données signifiante avec de bonnes performances.

Ci-dessous un tableau récapitulatif des différentes bases de données réalisées :

Tableau 3.1 : Récapitulatif des bases de données réalisées.

Bases de données	Lieu des enregistrements	Distance entre microphones (m)	Locuteurs masculins	Locuteurs féminins	Parole préenregistrée	Nombre de scénarios	Durée moyenne des scénarios (s)	Interlocuteurs par scénario	Positions fixes
BD 8	-laboratoire non vide. -Non sourd. -De: 5m x 3m	**4,20**	**3**	**2**	**0**	**27**	**7**	**1 ou 2**	-Gauche -Milieu_G -Milieu_D -Droite
BD 9	-chambre non vide. -mur couverts. -De: 3,8m x 3,4m	**2**	**1**	**1**	**1**	**15**	**22**	**1 ou 2**	-Gauche -Milieu -Droite
BD 10	-chambre non vide. -Non sourde. -De: 3,8m x 3,4m	**1**	**1**	**1**	**1**	**19**	**23**	**1 ou 2**	-Gauche -Milieu -Droite
BD 11	-chambre non vide. -Non sourde. -De: 3,8m x 3,4m	**1**	**2**	**2**	**1**	**6**	**94**	**1, 2 ou 3**	-Gauche -Milieu -Droite

- **Observation des résultats**

Les résultats de localisation des locuteurs sur les différentes bases de données réalisées sont illustres dans le tableau 3.2 suivant:

Tableau 3.2 : Résultats de localisation des locuteurs

Base de données	Résultats de localisation des locuteurs
BD 8	Echec total
BD 9	Faibles performances
BD 10	Bonnes performances, base de données non signifiante
BD 11	Bonnes performances, base de données signifiante

En examinant ces résultats on peut tirer les conclusions suivantes :

- Avec les deux bases de données BD8 et BD9, nous avons obtenus des très mauvais résultats qui sont dû probablement à l'effet d'écho.

- Dans la base de données BD10, le locuteur est localisé correctement, mais le nombre réduit des locuteurs (deux locuteurs) a fait que la base de données ne soit signifiante.

- Nous avons gardés la dernière base de données (BD11), car la localisation du locuteur est faite correctement et elle contient un nombre signifiant de locuteurs (3 locuteurs) et une variété de scénarios (6 scénarios).

III.3- Conception de la base de données BD11

III.3.1- Conditions de réalisation

D'après le tableau 3.1,cette expérience porte les données suivantes :

- La distance entre les deux microphones est : 1m.
- Lieu des enregistrements : chambre non vide, non sourde de 3,8m x 3,4m.
- Nombre de locuteurs masculins : 2.
- Nombre de locuteurs féminins : 2.
- Parole enregistrée : 1.
- Nombre de scenarios est : 6 scenarios.
- Nombre d'interlocuteurs par scenario : 1,2 ou 3 personnes différentes.
- Durée moyenne des scénarios est : 3mn 34s.
- Positions fixes : droite, milieu et gauche.

III.3.2- Scénarios de la base de données BD11

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 un locuteur masculin et un locuteur féminin.
- La troisième catégorie contient un discours entre plusieurs interlocuteurs (masculins et féminins) qui prennent la parole alternativement dans une manière naturelle.

Voici dans la figure 3.2 quelques images montrant certains scénarios de notre BD11 :



a)locuteur à gauche b) locuteur au milieu c) locuteur à droite



d) un locuteur à gauche et un à droite e) un locuteur à gauche et un au milieu

Figure 3.2: Photos montrant des locuteurs dans des différentes positions.

III.3.3- Configurations de la base de données BD11

Deux configurations possibles peuvent être proposéespour les différents scénarios: Une configuration fixe et une configuration mobile.
Dans la configuration fixe le locuteur est assis au tour d'une table dans une des trois positions fixes : gauche, milieu ou droite. Cependant, dans la configuration mobile, le locuteur se déplace lentement d'une position vers l'autre.

Dans notre base de données, on va étudier juste le cas de la configuration fixe comme illustrée dans la figure 3.3 :

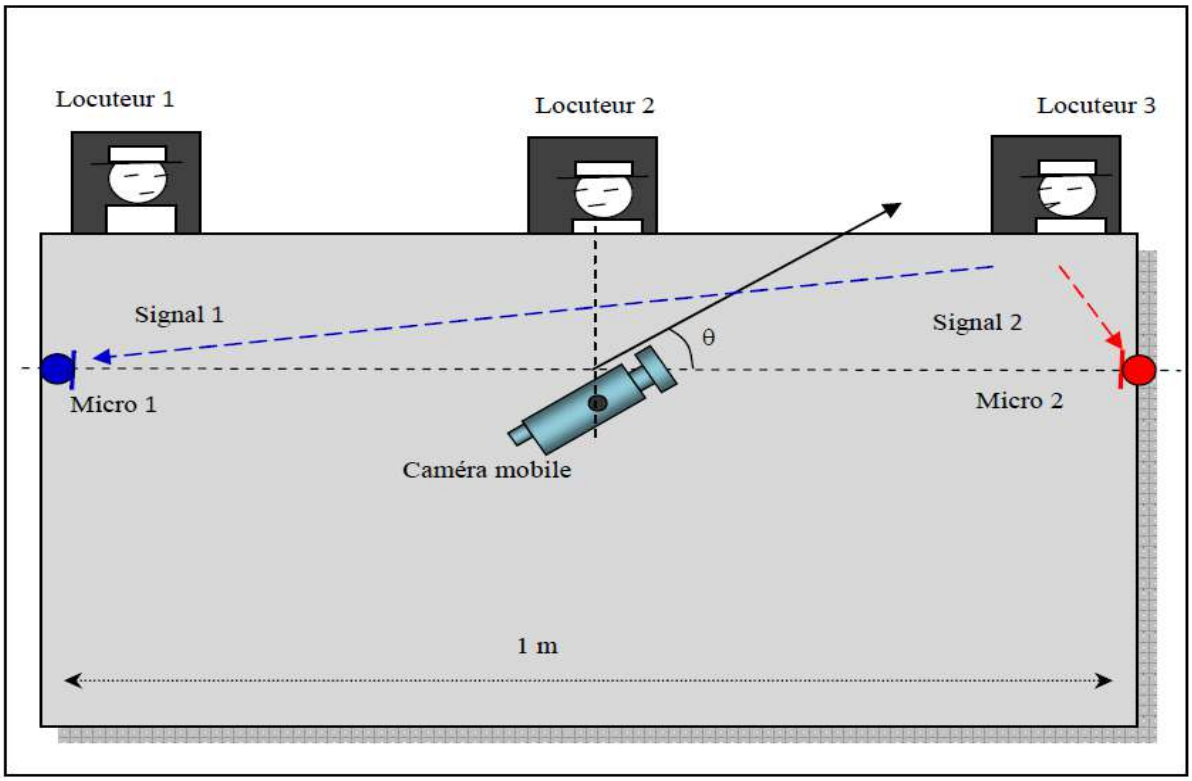


Figure3.3: Schéma synoptique de la configuration fixe.

III.2.4. Discours enregistrés

On a utilisés une variété de discours pendant les enregistrements audio de la base de donnéestels que : des sons phonémiques, des paragraphes, des discours et des conversations de manière naturelle.

III.4- Dispositifs matériels et logiciels utilisés

III.4.1- Schéma synoptique du système

Dans notre système, On distingue trois blocs essentiels: bloc d'acquisition, bloc de traitement et de commande et bloc d'exécution.

Le schéma synoptique du système automatique réalisé est illustré dans la figure 3.4:

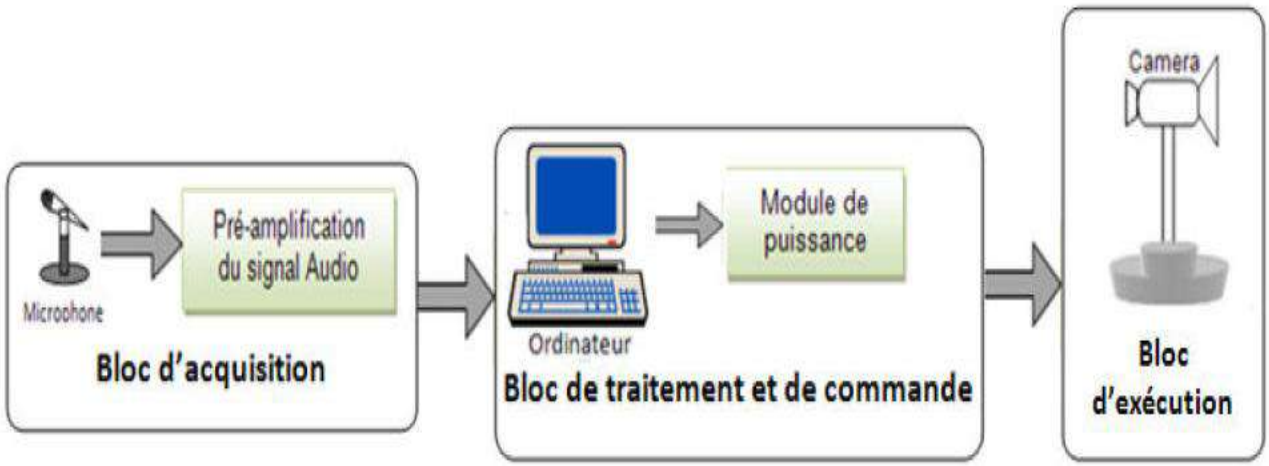


Figure 3.4: Schéma synoptique du système automatique réalisé.

III.4.2- Description des différents blocs du système

a) Bloc d'acquisition

Le bloc d'acquisition est constitué de deux microphones et d'un pré-ampli audio.
Le rôle de ce bloc estd'acquérir les signaux audio et de les amplifier avant qu'ils soient introduits au micro-ordinateur.

a.1) Microphones

Les signaux audio constituant la base de données BD11 sont perçus en utilisant deux microphones. Le premier est positionné à droite et le deuxième à gauche.



a) Microphone gauche b) Microphone droit

Figure3.5 : Photo des microphones utilisés dans les enregistrements.

Ces deux microphones sont les éléments du capteur les plus importants, plus le micro est performant, plus les mesures réalisées sur ce dernier sont précises. Ils possèdent les caractéristiques suivantes :

- Model dynamique unidirectionnel (cardioïdes).
- Portée de 2m maximale.
- Une réponsefréquentielle de 50Hz a 15000Hz.
- Une impédance de 150 Ohm.

a.2) Préampli audio

Les signaux audio enregistrés sont introduit par la suite dans unpré-amplificateur audio afin de les amplifier.
Le préampli utilisé est un préampli double voies, parfaitement adapté pour les microphones stéréo.
Il est basé sur l'emploi d'un circuit intégré de type NE5532. Ce circuit intégré comporte deux amplificateurs opérationnels au sein d'un mêmeboitier de huit broches, et qui présente de bonnes performances, notamment en termes de bruit. Il possède les caractéristiques suivantes:

- Impédance d'entrée : Environ 47KΩ pour chaque voie.
- Tension d'alimentation : +12 V.
- Gain du montage : 33dB.

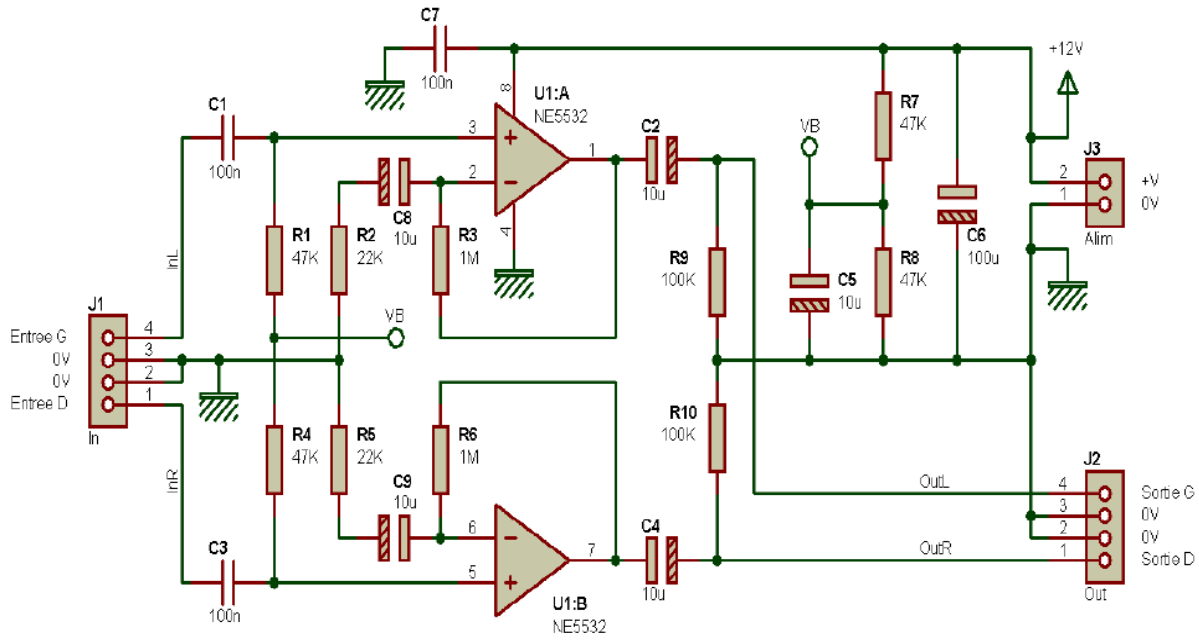


Figure 3.6: Schéma électrique du Pré-ampli audio utilisé.

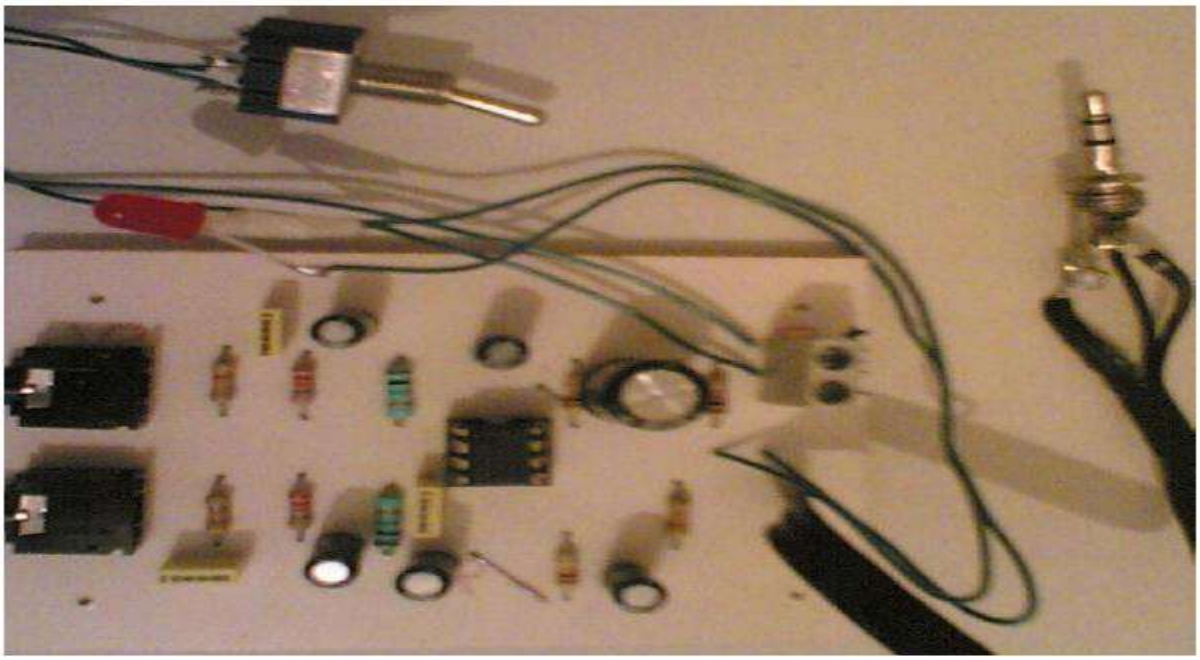


Figure3.7: Pré-ampli audio utilisé.

b) Bloc de traitement et de commande

Le rôle principal de ce bloc est de traiter les signaux audio enregistrés et de générer les signaux de commande adéquats. Il est constitué d'une unité de traitement et d'un circuit de puissance.

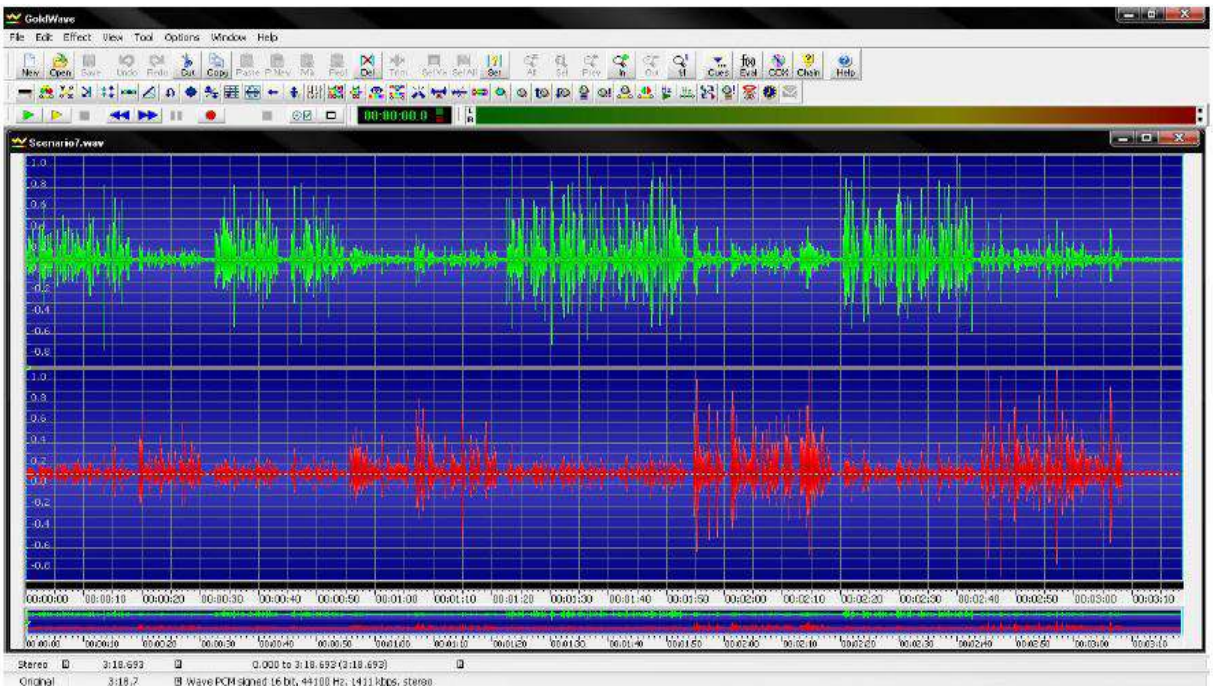
b.1) Unité de traitement

Elletraite les signaux perçus par les deux microphones (droit et gauche) et assure la commande du système via le port parallèle.
Cette unité de traitement ce n'est qu'un micro-ordinateur équipé de deux logiciels évolués :

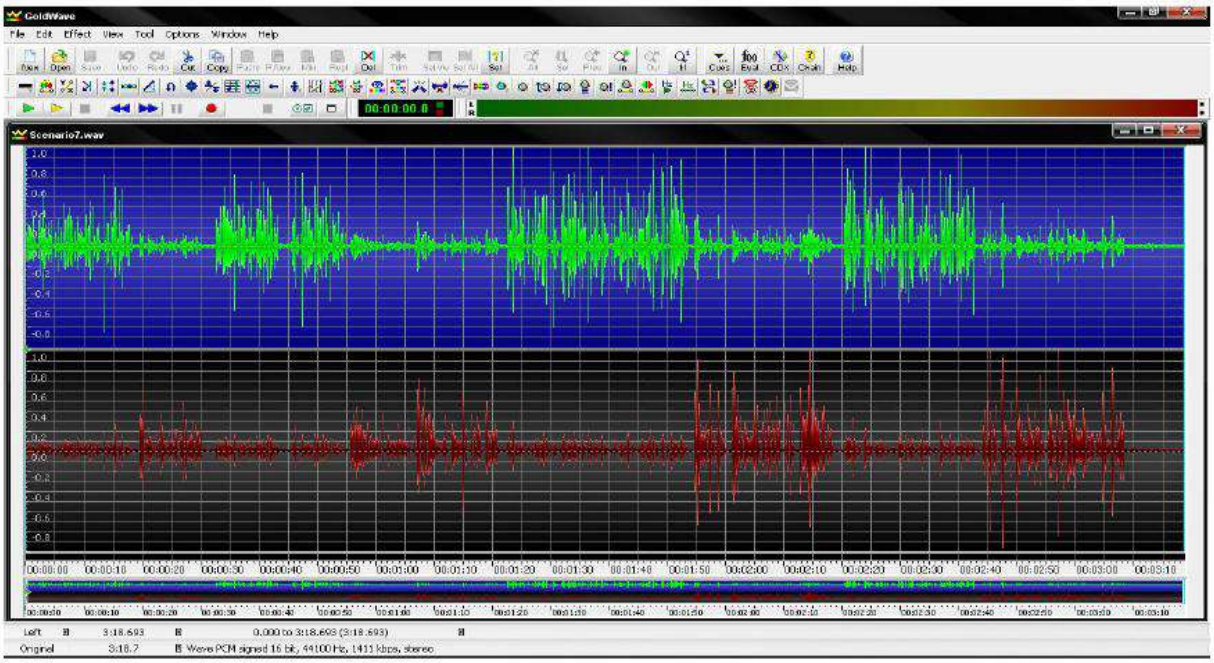
- Le Gold Wave.
- Le Matlab.

➢ **Pré-traitement audio avec le logiciel Gold Wave**

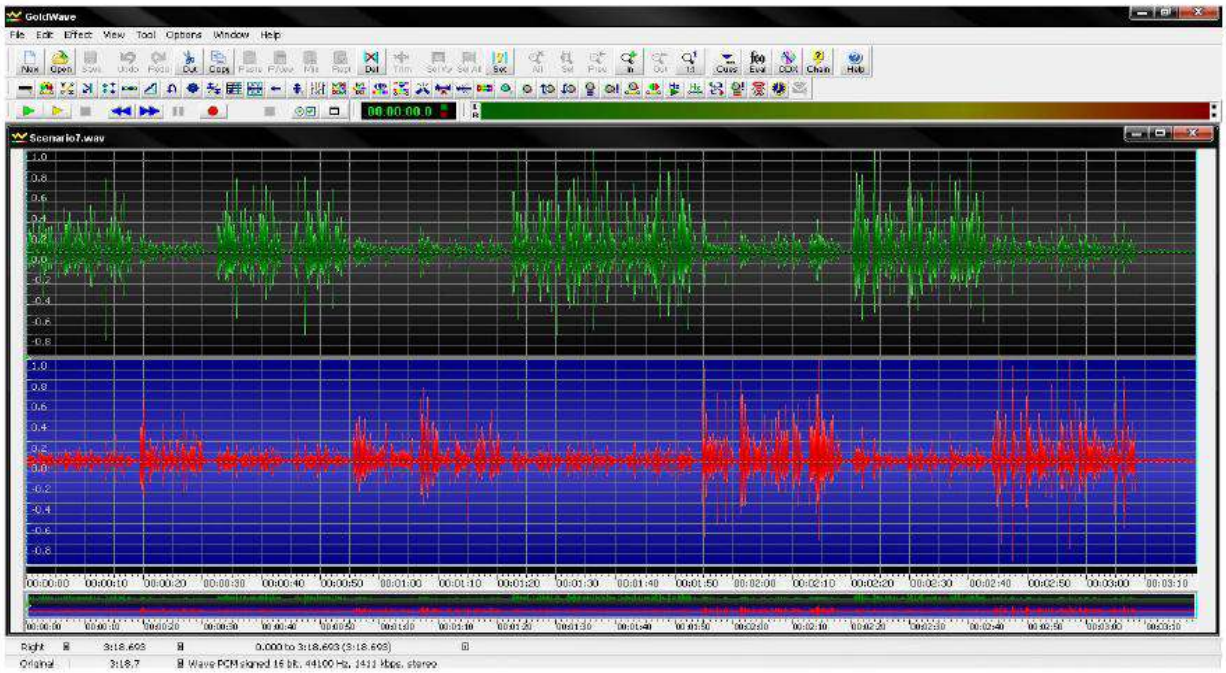
La dernière étape de préparation des signaux audio, qui sont des signaux stéréo, avant leur utilisation par notre système est le pré-traitement. Ce pré-traitement se fait par l'extraction des signaux de la voie gauche et de la voie droite du signal stéréo et les enregistrés sous des noms comportant l'indication de la voie concernée. Pour ce faire on a utilisé un logiciel appelé GoldWave.
Un exemple de ce pré-traitement est illustré par les photos présentées dans la figure suivante :



a) Signal stéréo (voie gauche et voie droite).



b) Sélection de la voie gauche.



c) Sélection de la voie droite.

Figure 3.8: Photo montrant le pré-traitement en utilisant le Gold Wave.

- **Traitement audio avec le logiciel Matlab**

Matlab est l'outil de référence pour la simulation numérique, notamment en ce qui concerne l'automatique et la commande. Il permet, de manière plus générale, de résoudre une grande diversité deproblèmes de simulation, dans des domaines aussi variés que le traitement du signal, les statistiques ou la vision.
C'est sur Matlab qu'on a essayé de faire nos programmes et de tester les performances des différentes méthodes de localisation proposées.

b.2) Circuit de puissance

L'interfaçage entre notre systèmeautomatique et l'ordinateur via le port parallèle est assuré par un circuit de puissance.

c) Bloc d'exécution

Lerôle de ce bloc est d'assurer l'orientation d'une caméra vers la direction appropriée.
A l'aide du logiciel Matlab, l'acquisition des images peut être faite directement à partir de la caméra et la transmettre à la carte graphique du micro-ordinateur.

III.5- Conclusion

Dans ce chapitre, nous avons présenté les détails de conception de notre base de données BD11 ainsi que lesdispositifs matériels et logiciels utilisés.
Avec cette base de données, on a obtenu des résultats satisfaisants en utilisant les méthodes de localisation proposées dans le chapitre précédant.

IV.1- Introduction

Dans ce chapitre, on va exposer les résultats des expériences effectuées en utilisant différentes méthodes de localisation appliquées sur plusieurs scénarios de parole. le logiciel utilisé est le Matlab.

Ces résultats seront interprétés et comparés d'une manière adéquate afin d'extraire les performances de chaque méthode quant à la qualité des résultats de localisation.

IV.2- Etapes des expériences réalisées

IV.2.1- Configuration de la base de données BD11

Notre base de données BD11 constituée de 6 scénarios, contenant des locuteurs masculins et féminins qui parlent de manière les naturelle, ces derniers sont assis dans des positions fixes. les expériences de localisation consistent à détecter la position du locuteur actif (qui parle) et ceci par le biais de sa parole.

IV.2.2- Acquisition des signaux audio

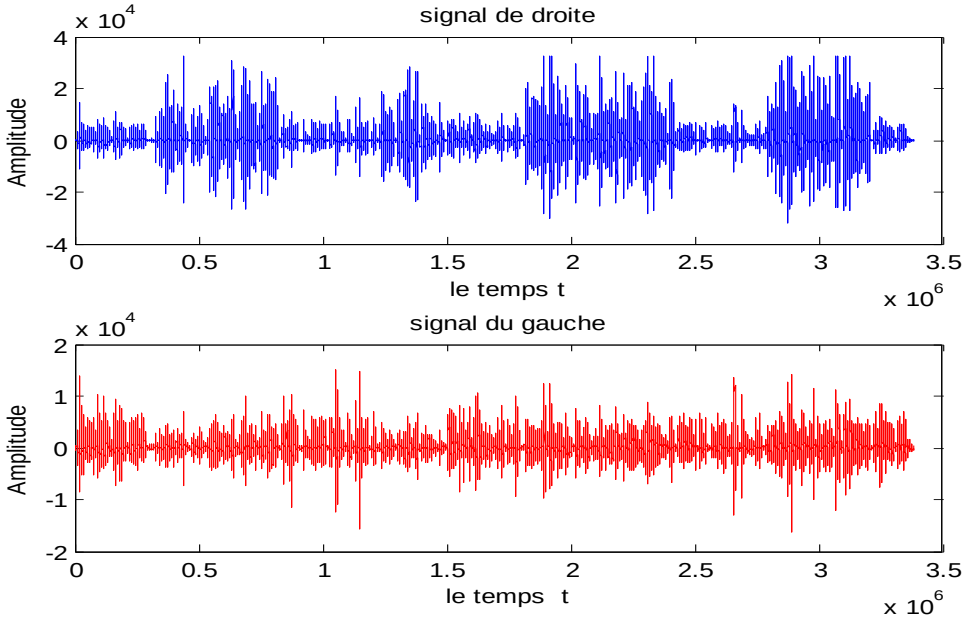
Les signaux audio sont enregistrés par deux microphones (gauche et droite), et amplifiés par un pré-ampli. les deux microphones sont positionné en opposition et séparés par une distance fixe (1mètre)

IV.2.3- Pré-traitement audio avec le logiciel Gold Wave

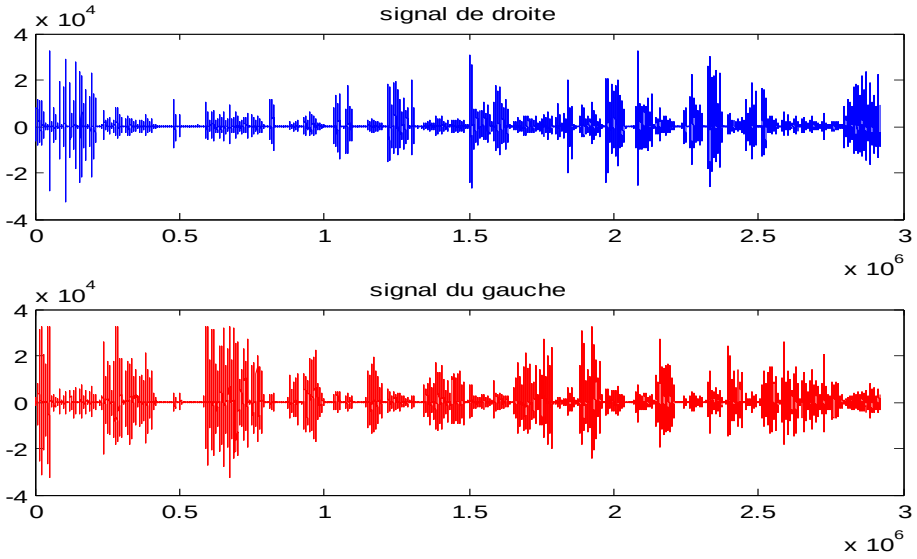
L'écoute est le premier moyen qui permet d'apprécier la qualité d'un signal, et de détecter les éventuels problèmes créées par le système de prise de son (Akchiche, 2007).

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

La figure 4.1est un exemple de lecture des signaux audio par le logiciel Gold Wave.



- a -



- b -

Figure 4.1: Visualisation des signaux audio.

a) Signaux issus d'une discussion entre deux locuteurs dont l'un est au milieu et l'autre est à gauche (scénario 1 de la BD11).

b) Signaux issus d'une discussion entre deux locuteurs dont l'un est à droite et l'autre est à gauche (scénario 5 de la BD11).

IV.2.4- Traitement audio avec le logiciel Matlab

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

- Méthode du différentiel d'énergie (ou en anglais Energy Differential Method **EDM**).
- Méthode de Generalized Cross-Correlation PHAse Transform (**GCC –PHAT**).

A fin d'obtenir une précision fine 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 représentés comme suit : la gauche (-1), le milieu (0) et la droite (1).

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

a) Méthode du différentiel d'énergie EDM

Avec cette méthode, la position est détectée selon la valeur de DE (différence d'énergie) qui est le différentiel de l'énergie logarithmique donné par la relation (2.4) (du chapitre 2).

Cette différence d'énergie DE sera comparée à un seuil de détection bien choisi expérimentalement.

b.1) Algorithme

L'algorithme suivi pour réaliser cette expérience est défini comme suit :

```
% la lecture des signaux issus respectivement du micro 'd' et du
micro 'g'

Lire (Sd)
Lire (Sg)

% fenêtrage des signaux

Décomposition des deux signaux de parole en fenêtres de taille 0.5 s
   Nombre de fenêtre = taille du signal/taille d'une fenêtre

%calcul de l'énergie des deux signaux et estimation de la position

 Pour   i=1:nombre de fenêtre

Calcul de l'énergie des deux signaux le long de la fenêtre

 Calculer (E_Sd (dB))
 Calculer (E_Sg (dB))

Calcul de différentiel d'énergie

 DE = E_Sd (dB) - E_Sg (dB)

Estimation de la position
     Si DE>seuil_max
                Position='droite'
     Sinon Si DE<seuil_min
                Position='gauche'
     Sinon
                Position='milieu'
     Fin Si;

Fin pour
```

b.2) Résultats des expériences réalisées

En appliquant l'algorithme précédent aux différents scénarios de la base de données BD11, on a obtenu les résultats illustrés dans les figures (4.2) à (4.7).

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

Scénario 1: Halim-Hakim_ (droite-milieu)

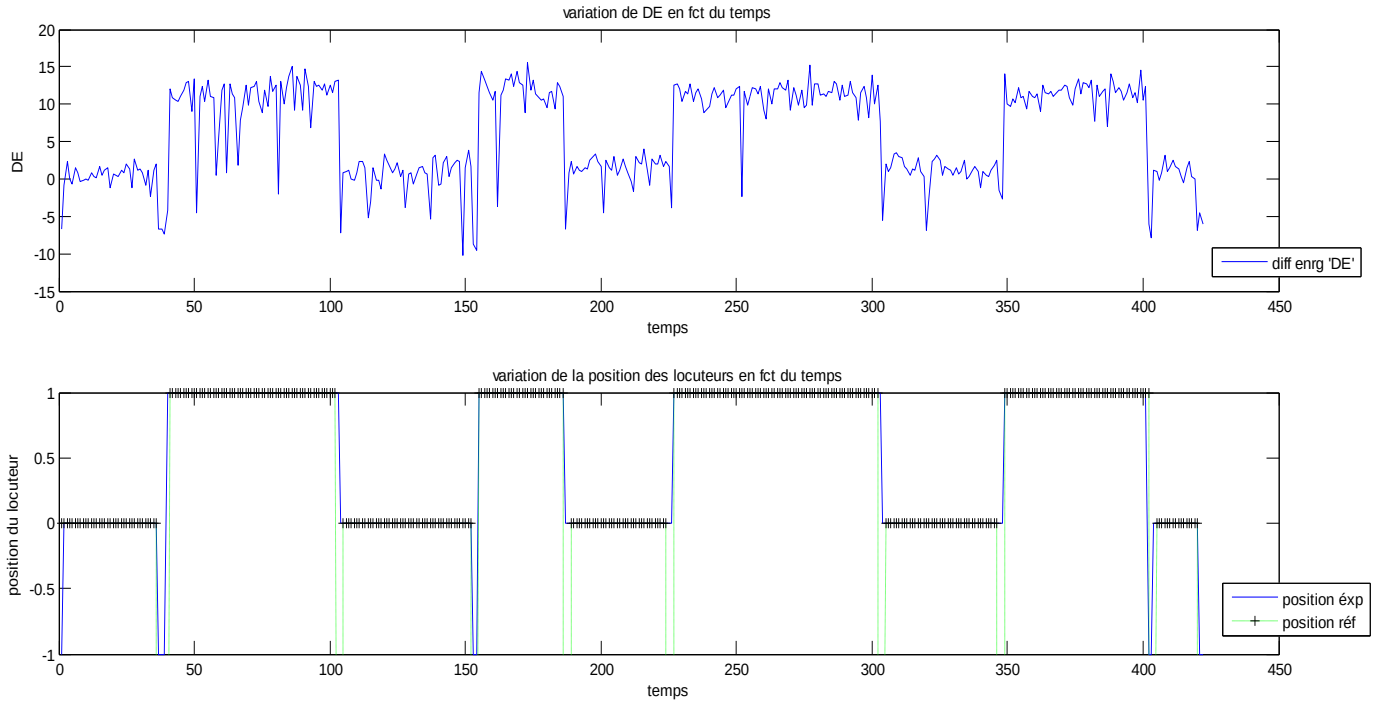


Figure 4.2 : Résultats de localisation obtenus par la méthode EDM pour le scénario 1.

Scénario 2: Halim-Siham_(milieu-droite)

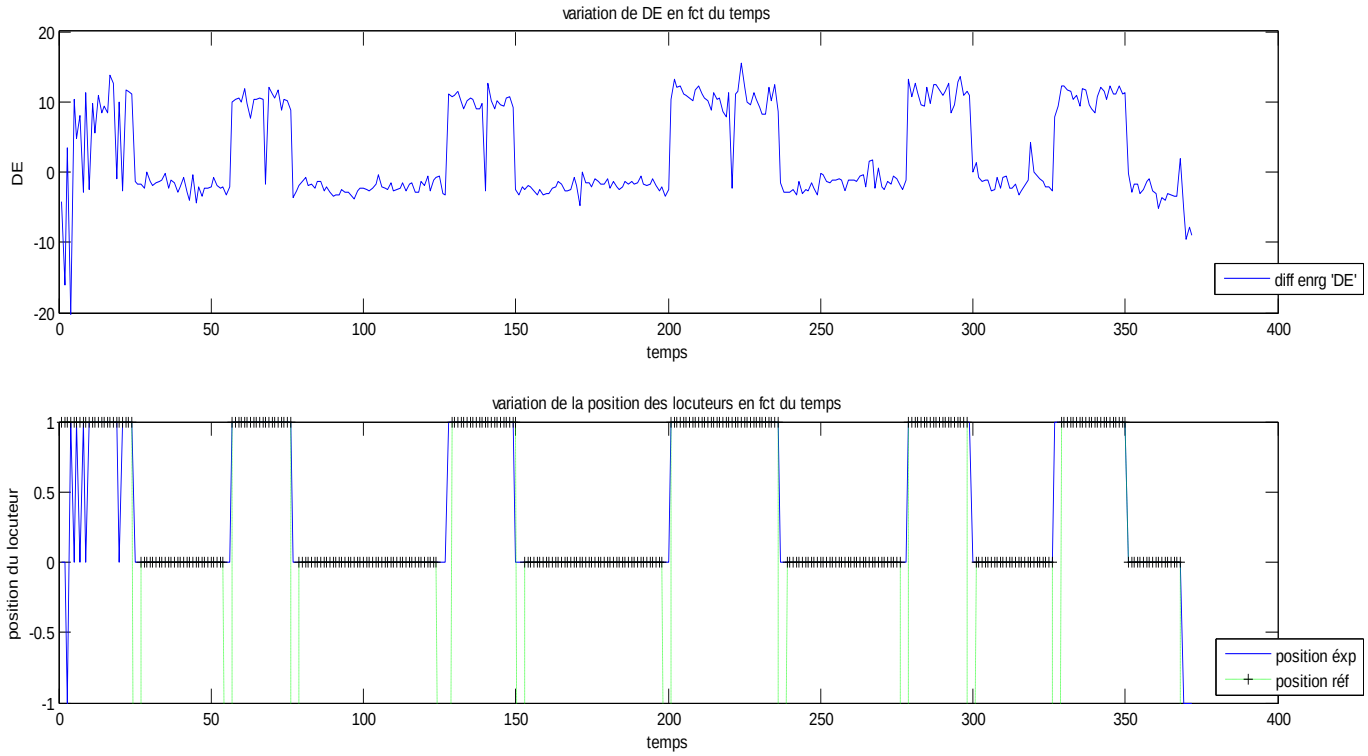


Figure 4.3 : Résultats de localisation obtenus par la méthode EDM pour le scénario 2.

Scénario 3: Nadia-Halim (gauche-milieu)

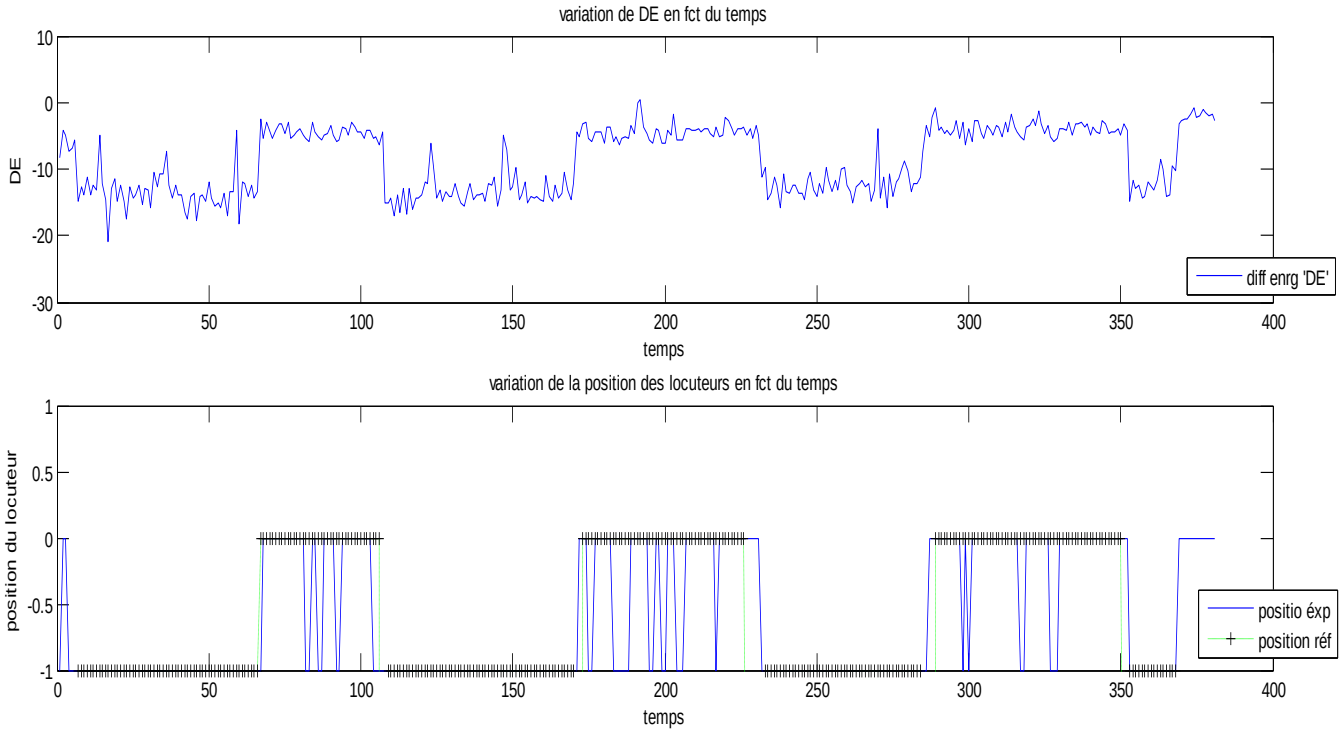


Figure 4.4 : Résultats de localisation obtenus par la méthode EDM pour le scénario 3.

Scénario 4 : Hakim-Halim_(gauche-droite)

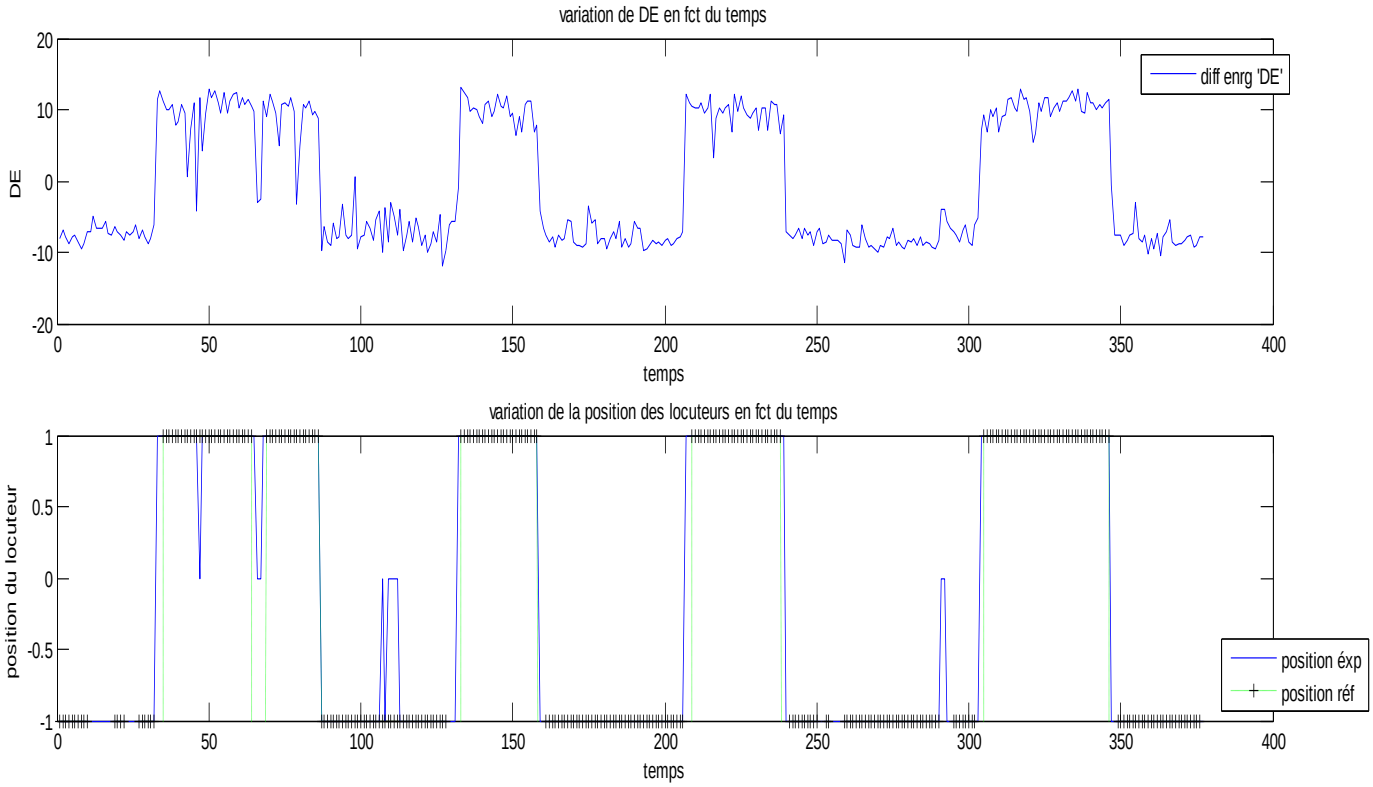


Figure 4.5 : Résultats de localisation obtenus par la méthode EDM pour le scénario 4.

Scénario 5:Halim-Siham_(gauche-droite)

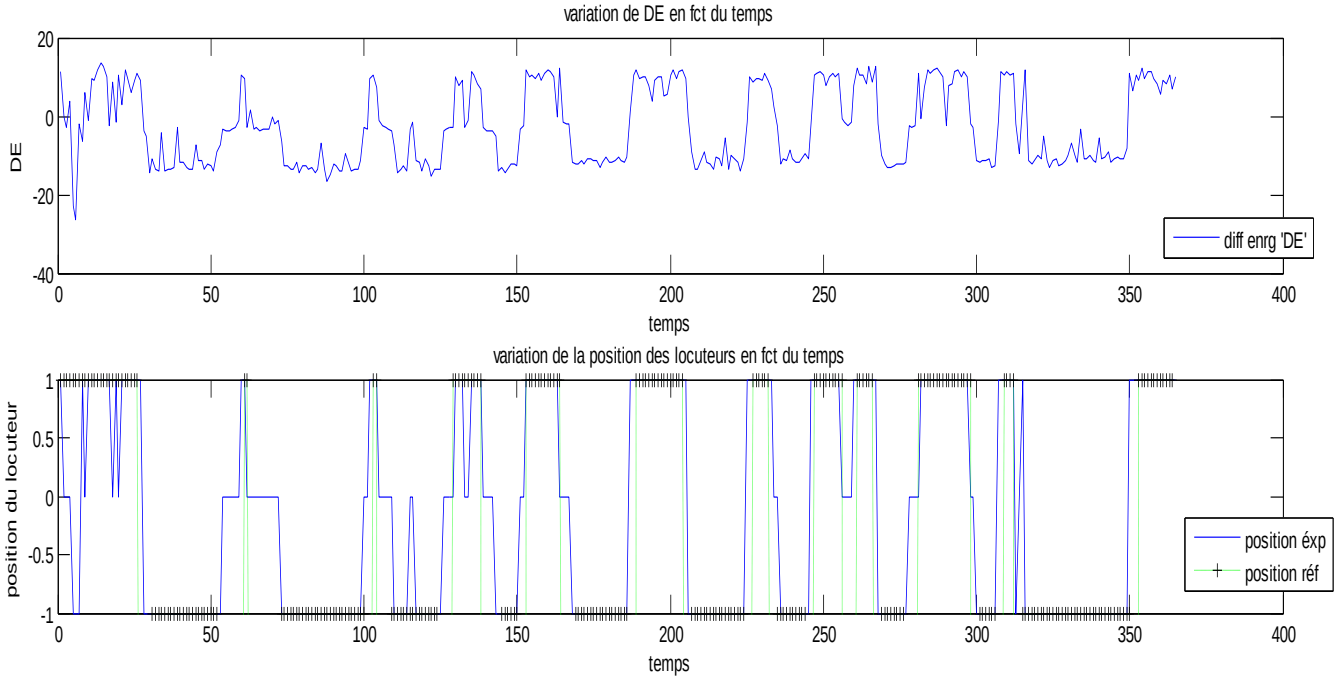


Figure 4.6 : Résultats de localisation obtenus par la méthode EDM pour le scénario 5.

Scénario 6:Nadia-Halim_ (gauche-droite)

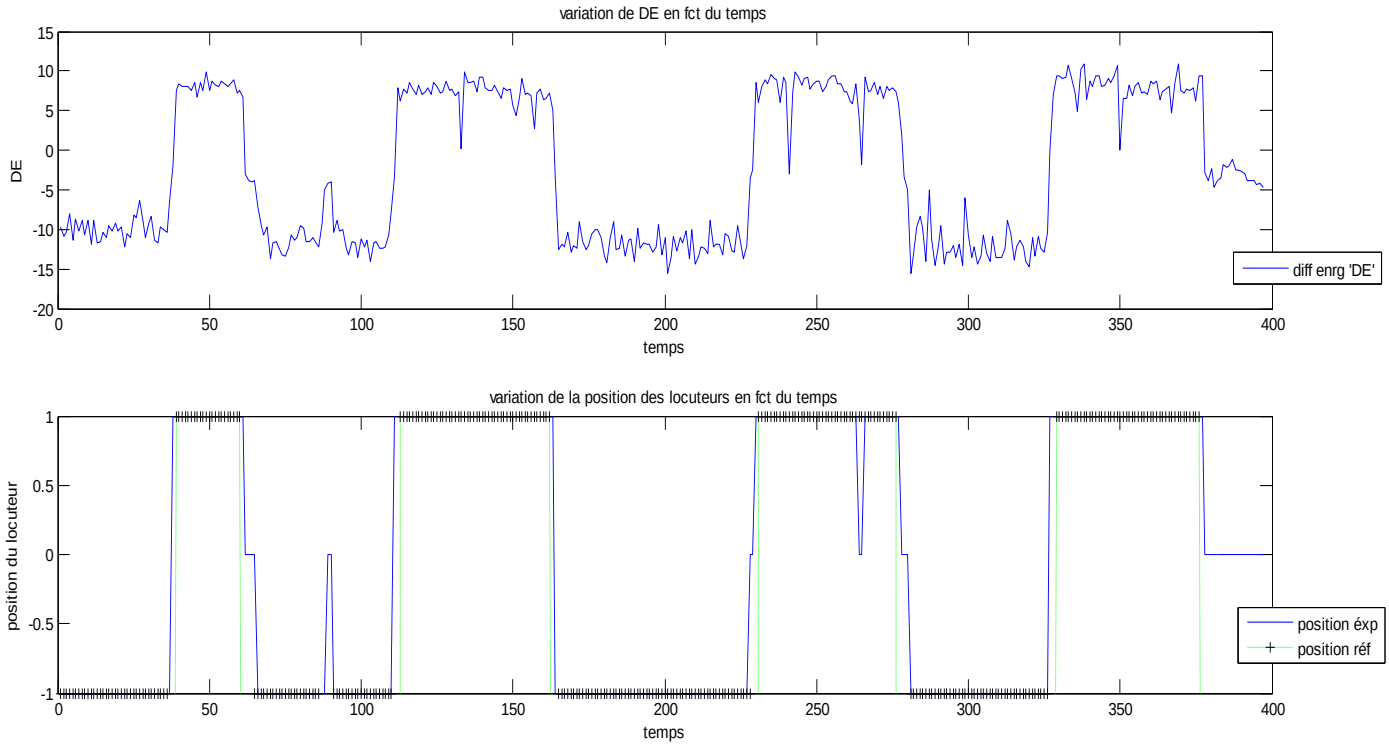


Figure 4.7 : Résultats de localisation obtenus par la méthode EDM pour le scénario 6.

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

Pour la méthode de la **GCC-PHAT** : on a utilisé la formule (2.27) du chapitre 2, qui nous permet de calculer le retard ($\boldsymbol{\tau}$) 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) donnée par la relation (2.8) du chapitre 2.
Cette distance sera comparée un seuil de détection bien choisi expérimentalement.

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

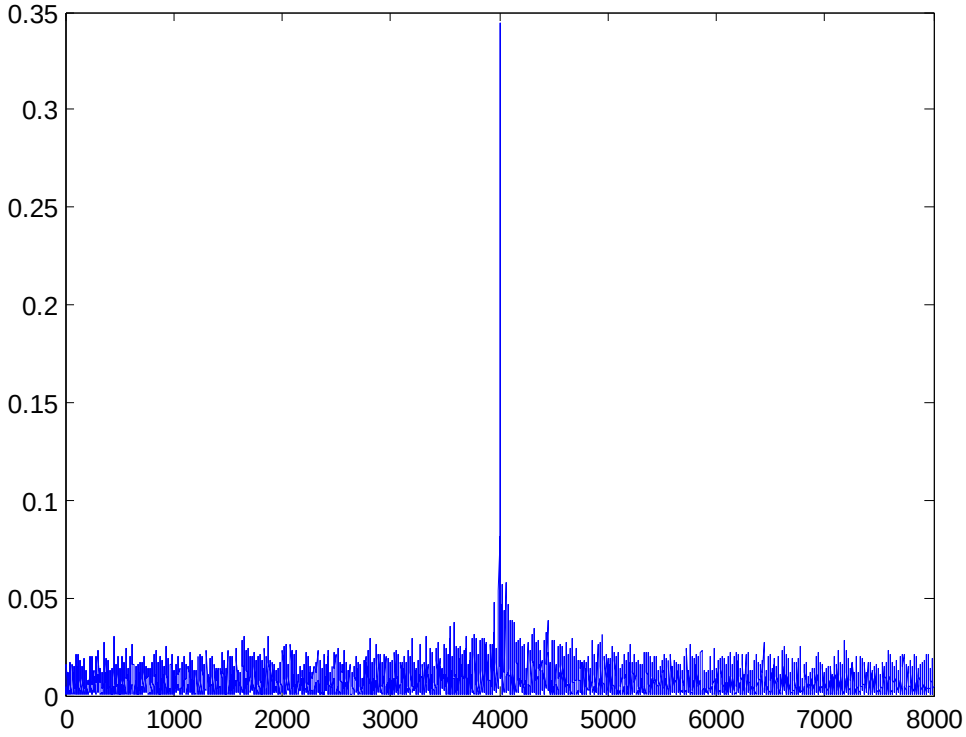


Figure 4.8 : exemple de localisation d'un locuteur au milieu (scénario 1)

- Le pic donne un retard $\tau = 2{,}5\ e^{-4}$ s (4.1)
  La distance **d** est : d=0,5426 m (0,4<d<0,6) (4.2)

Donc : le locuteur qui parle est au milieu.

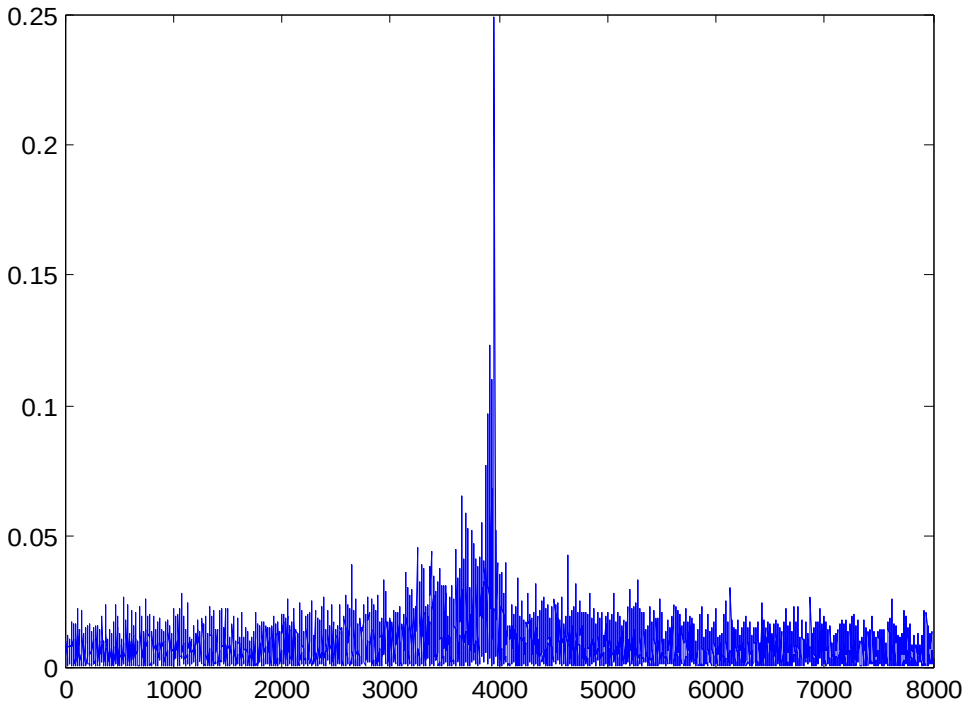


Figure 4.9 : exemple de localisation d'un locuteur à droite (scénario 1)

- Le pic donne un retard $\tau = -0.0024s$ (4.2)
  La distance **d** est : d=0,092 m (d<0,4) (4.3)

Donc : le locuteur qui parle est à droite.

c.1) Algorithme

Voici l'algorithme suivi pour réaliser cette expérience :

```
% la lecture des signaux issus respectivement du micro 'd' et du
micro 'g'

Lire (Sd)
Lire (Sg)

% fenêtrage des signaux

Décomposition des deux signaux de parole en fenêtres de taille 0.5 s
   Nombre de fenêtre=taille du signal/taille d'une fenêtre

%calcul du retard (τ) et estimation de la position

 Pour   i=1:nombre de fenêtre

Calcul de la GCC-PHAT sur chaque fenêtre
```

$$GCC - PHAT = ifft\left(\frac{A.B^*}{|A|.|B|}\right)$$

```
Détection de la position du maximum, donc (τ)
```

$$\tau = argmax\,(GCC - PHAT)$$

```
          Calcul de la distance entre les locuteurs et les microphones
```

$$d=(L+\tau*Vs)/2;$$

```
Estimation de la position
     Si d>seuil_max
                Position='droite'
     Sinon Si d<seuil_min
                Position='gauche'
     Sinon
                Position='milieu'
     Fin Si;
Fin pour
```

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

Scénario 1:Halim-Hakim_ (droite-milieu)

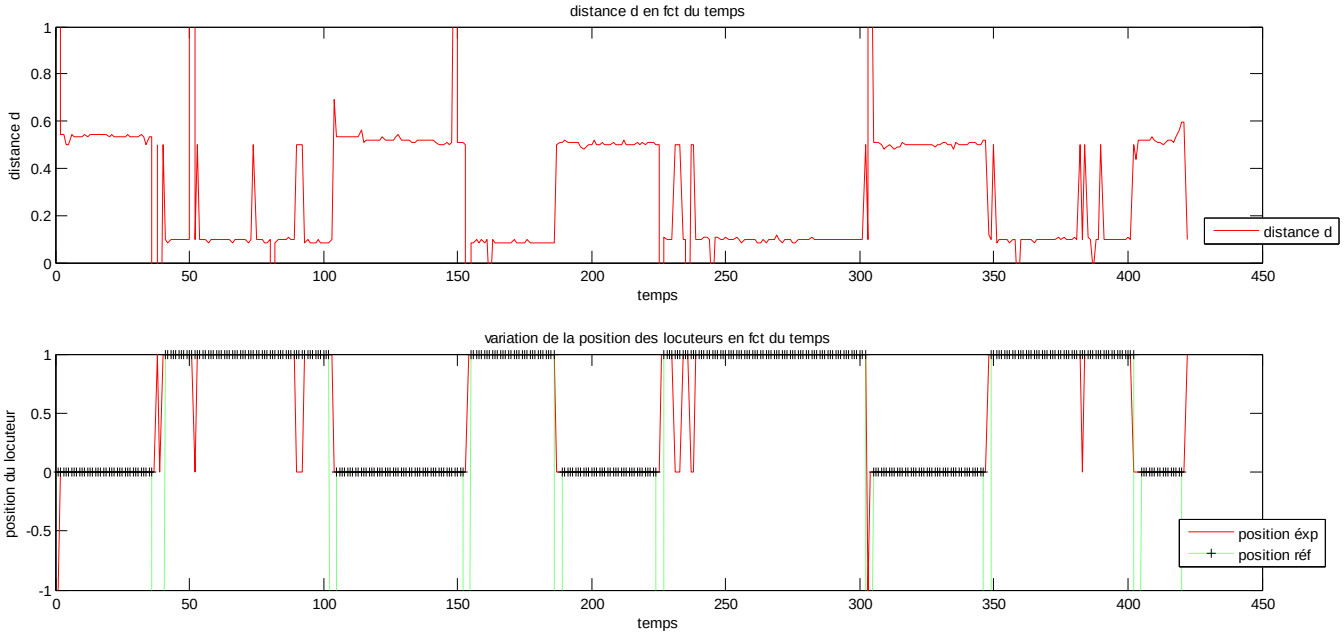


Figure 4.10: Résultats de localisation obtenus par la méthode GCC-PHAT pour le scénario 1.

Scénario 2:Halim-Siham_(milieu-droite)

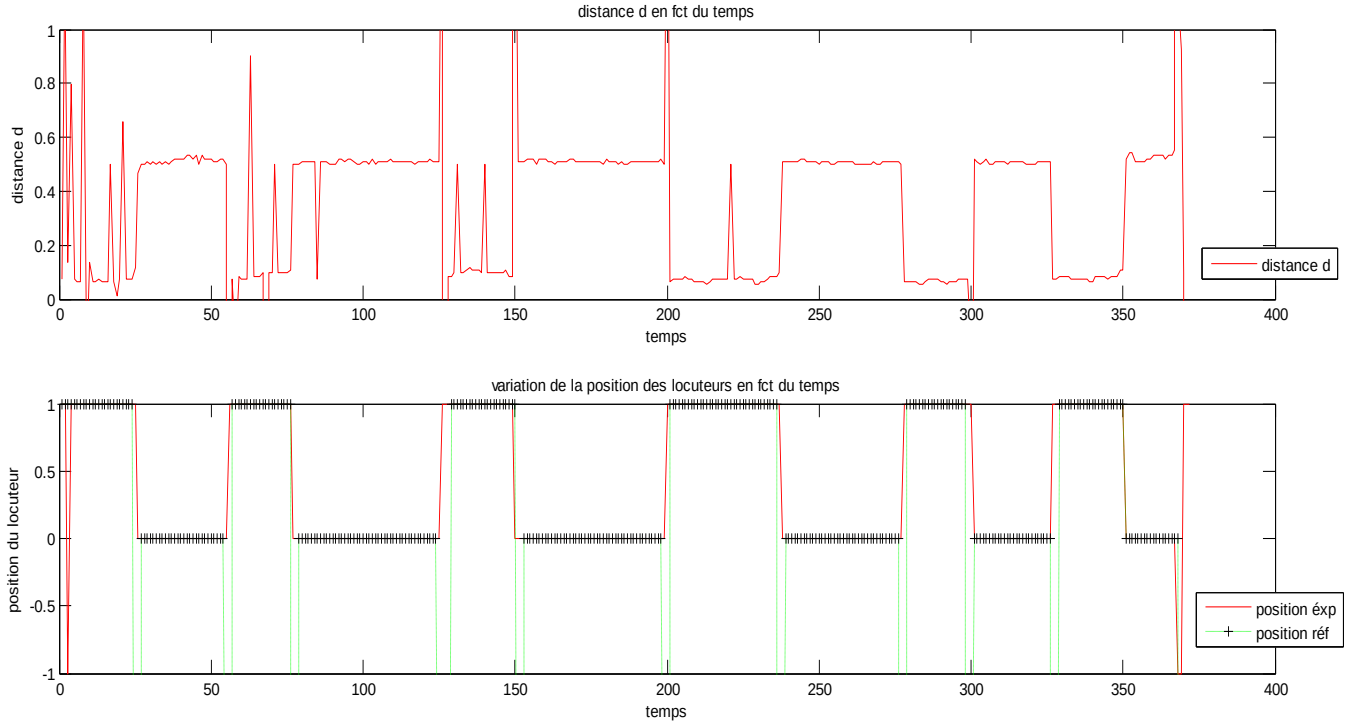


Figure 4.11 : Résultats de localisation obtenus par la méthode GCC-PHAT pour le scénario 2.

Scénario 3:Nadia-Halim_ (gauche-milieu)

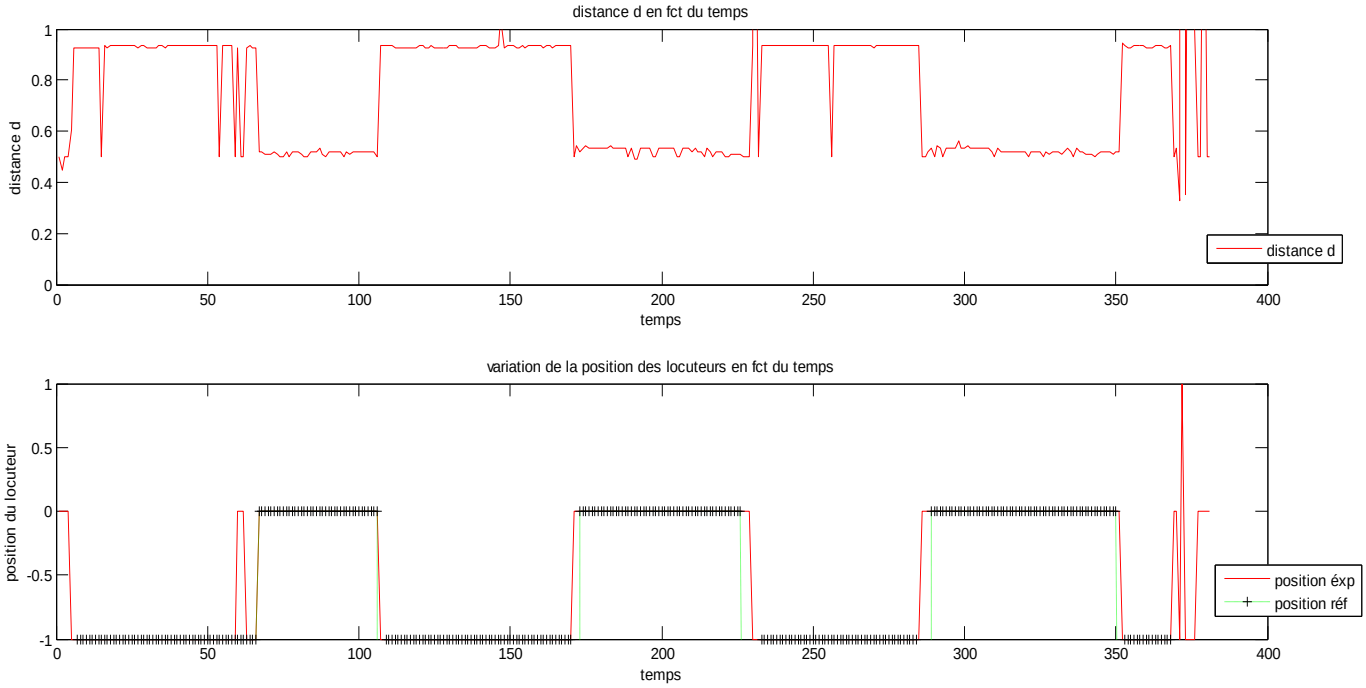


Figure 4.12 : Résultats de localisation obtenus par la méthode GCC-PHAT pour le scénario 3.

Scénario 4:Hakim-Halim_(gauche-droite)

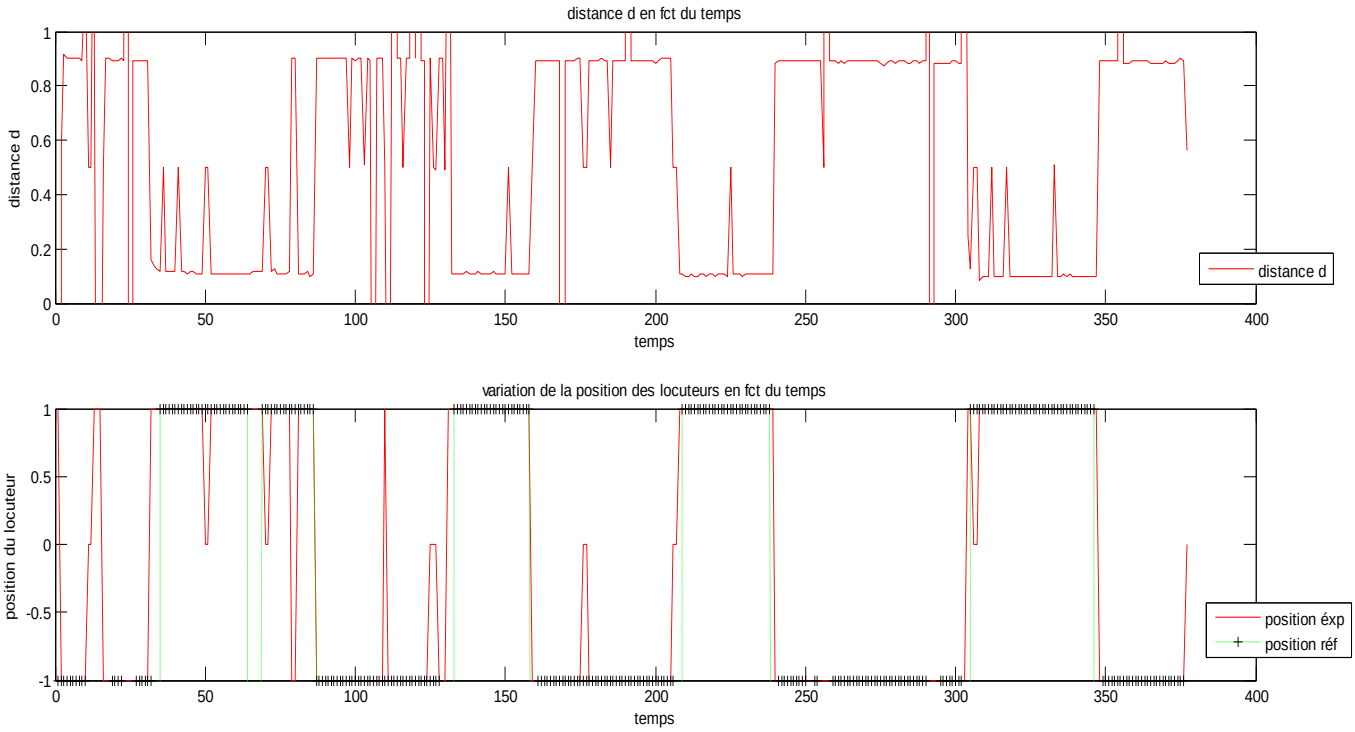


Figure 4.13 : Résultats de localisation obtenus par la méthode GCC-PHAT pour le scénario 4.

Scénario 5:Halim-Siham_(gauche-droite)

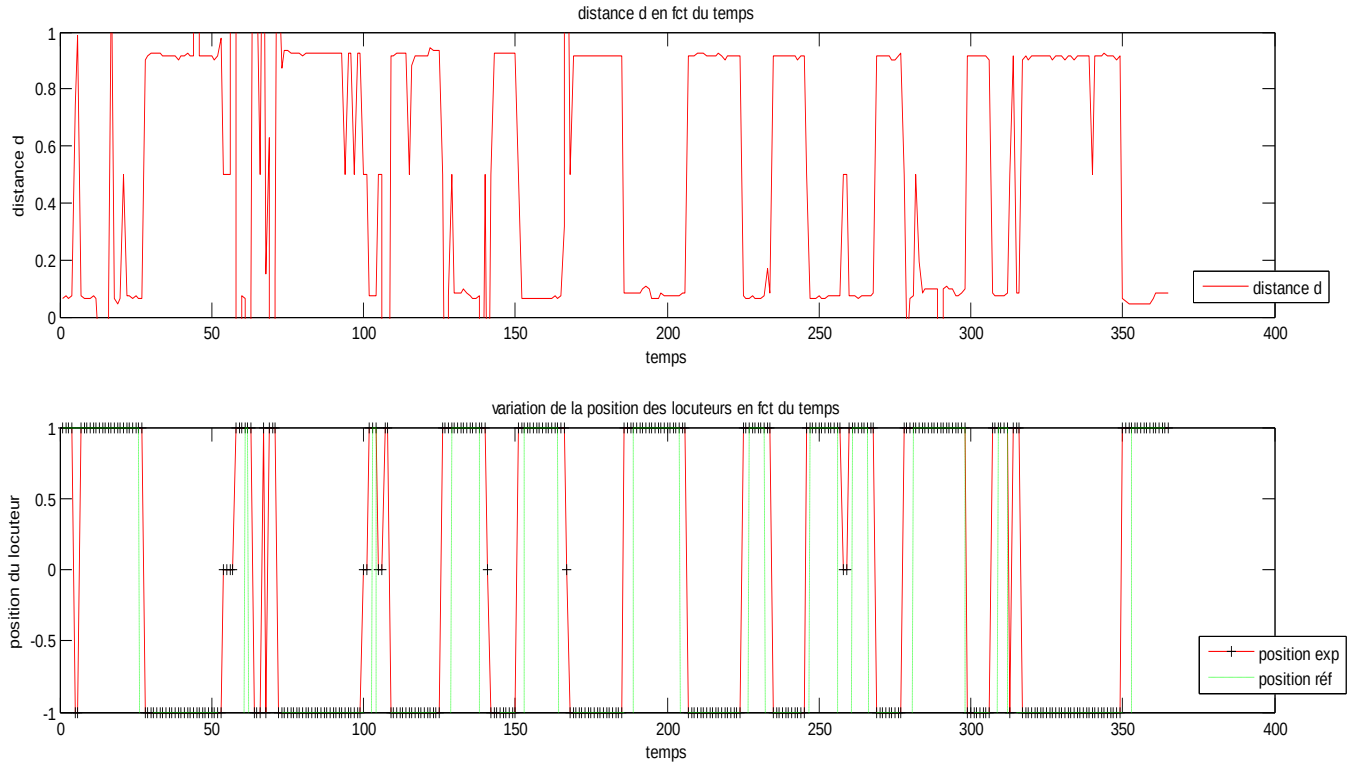


Figure 4.14 : Résultats de localisation obtenus par la méthode GCC-PHAT pour le scénario 5.

Scénario 6:Nadia-Halim_(gauche-droite)

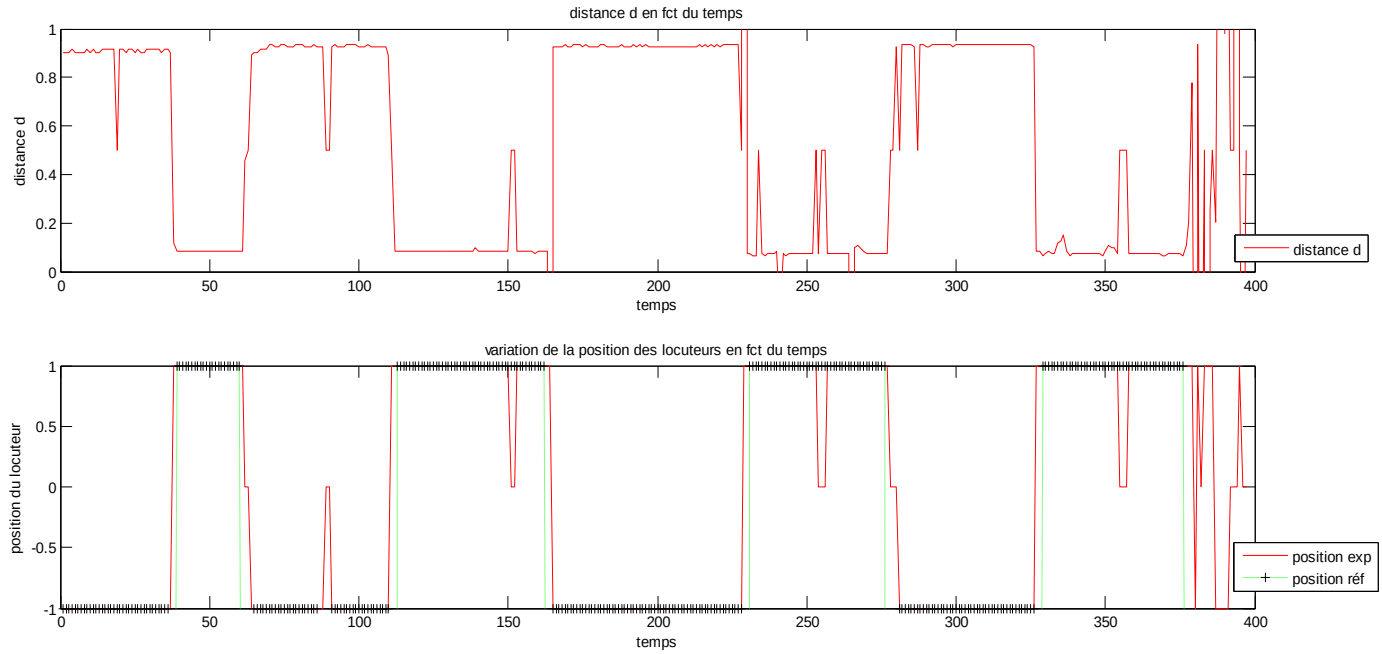


Figure 4.15 : Résultats de localisation obtenus par la méthode GCC-PHAT pour le scénario 6.

IV.3- Statistiques des résultats obtenus

La simulation avec Matlab a donné les résultats illustrés dans le tableau ci-dessous:

Tableau 4.1: Taux de précision de localisation de la position en % par scénario.

techniques / scénarios	EDM	GCC-PHAT
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**

Avec les résultats, on a tracé l'histogramme suivant:

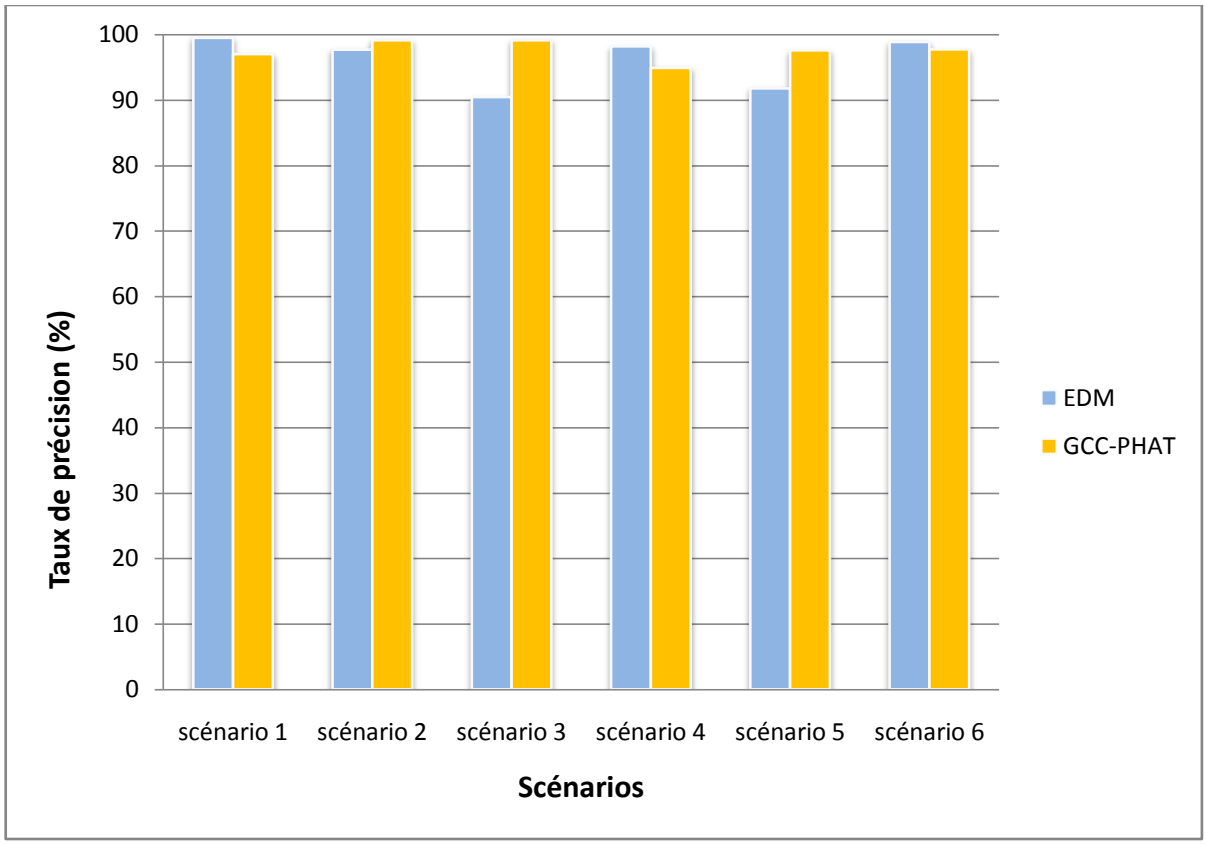


Figure 4.16: Taux de précision de localisation de la position en % par scénario.

Commentaire

- La méthode **EDM** présente une variation de taux de précision de la localisation entre 90,46 % et 99,5 %. Ces valeurs changent d'un scénario à un autre.

- La méthode **GCC-PHAT** présente une variation de taux de précision de la localisation entre 94,91 % et 99,13 %. Ces valeurs changent d'un scénario à un autre.

Pour pouvoir comparer les performances des deux méthodes, il est préférable de calculer la moyenne de taux de précision de chacune d'elle.

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

Tableau 4.2: La moyenne des taux de précision par méthode.

EDM	GCC-PHAT
96 ,03 %	**97,59 %**

Les résultats obtenus sous forme d'histogramme est dans la figure suivante :

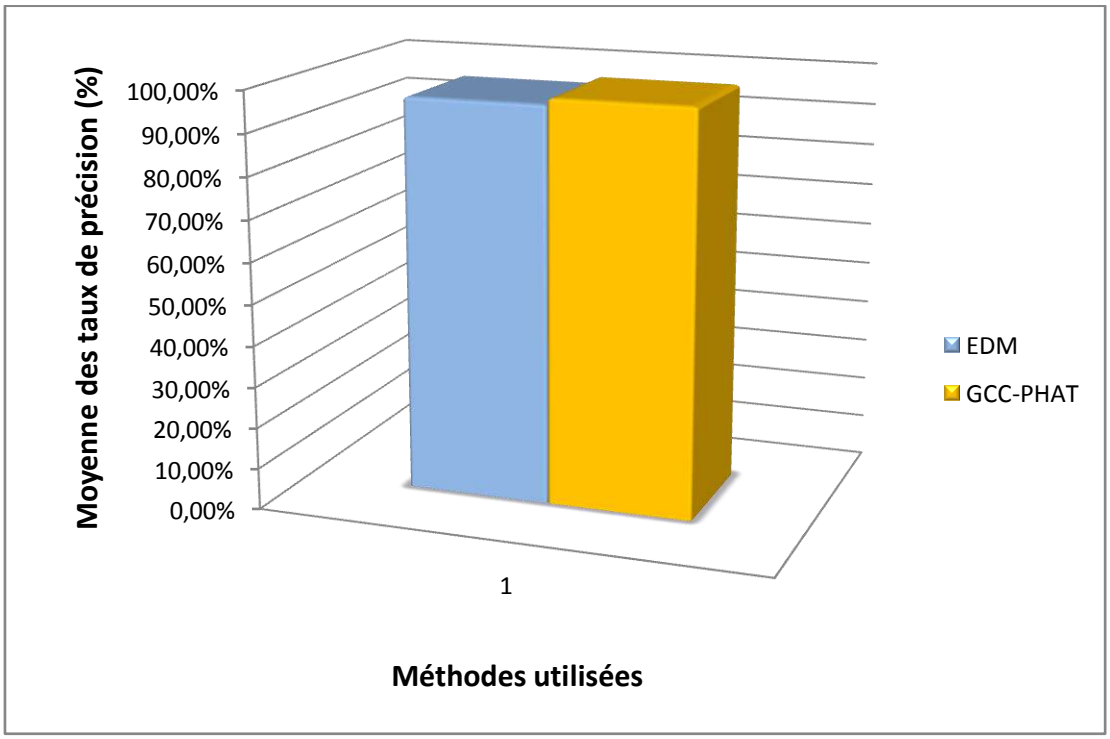


Figure 4.17: La moyenne des taux de précision pour chaque méthode.

Commentaire :

Les deux méthodes **EDM** et **GCC-PHAT** donnent des résultats proches et satisfaisants concernant le taux de précision de la localisation mais la **GCC-PHAT (97,59 %)** était un peu plus meilleure car elle est très proche de l'objectif voulu**.**

IV.4- Conclusion

Dans ce chapitre, on a pu montrer l'efficacité des trois techniques utilisées dans la détection spatiale des locuteurs avec des signaux stéréophoniques en utilisant une paire de microphones.

Parmi ces techniques, la CCG-PHAT était la plus performante avec un taux de précision qui atteint les **97,59 %.**

Comme nous l'avons mentionné dans l'introduction de ce manuscrit, le but de notre travail est d'effectuer une étude sur la localisation des sources sonores présentée par des locuteurs qui parlent à l'aide des signaux stéréophoniques.
Notre application était destinée à la supervision automatique des enregistrements audiovisuels dans les meeting-rooms dans le but d'améliorer les performances des systèmes de détection du son.

Au début nous avons présenté les différentes méthodes de détection telle que: le différentiel d'énergie **(EDM)** et la Generalized cross-correlation phase transform (**GCC –PHAT**).

La plupart de ces méthodes a fait l'objet de plusieurs études comparatives qui ont montré les avantages des unes et les inconvénients des autres.

Et pour pouvoir juger d'une manière adéquate nos résultats, plusieurs scénarios ont été prise en compte, parfois, on change les locuteurs (masculins ou féminins) et d'autre fois, on change leurs positions (gauche, milieu ou droite).

Discussion des résultats obtenus :

Les résultats expérimentaux obtenus ont montrés :

- Une bonne précision de localisation oùla moyenne des deux techniques dedétection atteint environ les **96,81 %**de bonne détection.

- Une meilleure précision de détection pour la méthode basée sur la GCC-PHATqui atteint environ les **97,59 %**surcelle qui utilise le différentiel d'énergieEDM.

Ces résultats sont obtenus pour des positions fixes et en utilisant un nombre limité de locuteurs (3 personnes seulement) ainsi on a traité uniquement deux techniques de localisation (EDM et GCC-PHAT) mais il sera intéressant d'augmenter le nombre de locuteur (gauche, droite, milieu, milieu gauche et milieu droite) et appliquer ces méthodes sur des locuteurs mobiles ainsi d'utiliser d'autres techniques de localisation.
De plus, les erreurs de localisation peuvent être réduites d'avantages si on renforce les conditions d'enregistrement audio par une réduction des échos (chambre sourde) ou par l'utilisation de microphones de meilleure directivité tel que les microphones matricielles (microphone arrays) qui nous offre la possibilité de localiser les locuteurs dans le plan (X,Y) et dans l'espace (X,Y,Z).

Ceci n'empêche pas de dire qu'à travers cette étude, nous estimons que le projet que nous avons mené nous a été fort utile de par les connaissances théoriques mais surtout pratiques que nous avons acquis.

- ***Les documents écrits :***

(Akchiche**, 2007)** Akchiche S.,"Identification du locuteur: Approche par quantification vectorielle", Mémoire d'ingéniorat en génie électrique, Ecole Militaire Polytechnique, 2007.

(Ait Saadi,**2005**) Ait Saadi N.,Takhedmit H., "Identification du locuteur en mode indépendant du texte", Mémoire de Doctorat en électronique, Ecole Nationale Polytechnique, El Harrach, Alger, Juin 2005.

(Benselama,**2007**) BenselamaZoubir-Abdeslem.,"Pathologie du Langage Parlé Arabe Cas des Sigmatismes Occlusifs et Constrictifs", Mémoire de Doctorat d'état en électronique, Ecole Nationale Polytechnique, El Harrach, Alger, Décembre 2007.

(Boite ,**1987**) Boite R., Kunt M.," Traité d'électricité : Traitement de la Parole", Presses Polytechniques Romandes, Lausanne, Suisse, 1987.

(Brutti ,**2008**) Brutti A, Omologo M., Svaizer P., "Comparaison between different sound source localization techniques based on a real data collection", IEEE HSCMA 2008.

(Hoang ,**2009**) Hoang Tran Huy Do., "Real-time SRP-PHAT source location: implementations on a large-aperture microphone array", Master of science in the division of engineering, Brown university, September, 2009.

(Khennouf,**2010**) Khennouf S., "Système automatique pour l'orientation de caméra mobile vers des cibles sonores", Mémoire de Magister en électronique: contrôle des processus et robotique, USTHB, Alger, 2010.

(Kunt,**1996**) Kunt M., "Traitement numérique des signaux". Edition technique et documentation, Lavoisier, 3ème édition, Lausanne, Suisse, 1996.

(Larousse,**1969**) Larousse P., Dictionnaire encyclopédique : Nouveau petit LAROUSSE, Paris, 1969, pp.719- 956.

(Margueron,**2000**) Margueron Xavier, "Capteur pour la localisation d'une source sonore", Département Génie Electrique et Informatique Industrielle 1: Electrotechnique et Electronique de Puissance, INRIA Rhône-Alpes, 25 Avril-30 Juin 2000.

Les sites web :

(Matlab7:programmingfondamentals,**2011**)
"Matlab7: programmingfondamentals".
http//www.mathworks.com
(Consulté le 19/06/2011 12:53)

(La *localisation* de la *source sonore*,**2012**) "La *localisation* de la *source sonore*"
http://cours.musique.umontreal.ca/MUS3321/MUS3321_SLIDES/MUS3321cours9_localisation.pdf
(Consulté le28/02/2012 14:36)

(*Localisation* d'un point d'émission d'une *source sonore* dans l'espace,**2012**) "*Localisation* d'un point d'émission d'une *source sonore* dans l'espace"
http //www.phaz.mc/edu/acoustique/localisation.html
(Consulté le22/02/2012 22:42)

(Pourquoi est-ce plus dur de localiser les sons aigus que les sons graves?,2012) "Pourquoi est-ce plus dur de localiser les sons aigus que les sons graves? "
http://philipe.boeuf.pagesperso -organe.fr/robert/physique/localisation-son.htm
(Consulté le 22/02/2012 22:36)

(Correlation,wikipédéa,**2012**) http://fr.wikipédia .org /wiki/ cor % c3 % a9latio_(statistiques).
(Consulté le 05/04/2012 14:18)

(Autocorrelation,wikipédéa,**2012**) http://fr.wikipédia .org /wiki/ autocorr % c3 % a9latio # traitement-du-signal.
(Consulté le 05/04/2012 14:23)

République Algérienne Démocratique et Populaire
Ministère de l'enseignement Supérieur et de la Recherche Scientifique
Université de Dr Yahia Fares -MEDEA-

Faculté d'Electronique et d'Informatique

Domaine Sciences Techniques

Mémoire de Master

Génie Electrique

Option: Systèmes de Télécommunications

Thème

DETECTION SPATIALE DES LOCUTEURS PAR DES SIGNAUX STEREOPHONIQUES.

Proposé et dirigé par :

Mme OUAMOUR Siham.

Mr SAYOUD Halim

Présenté par :

■ **M^{elle} DJEMAI Lynda.**

■ **M^{elle} SADI Hadjer.**

JUILLET 2012

Remerciements

*D'abord nous remercierons **DIEU** pour la force que nous a donné afin de terminer ce travail dans la tranquillité.*

Nos vifs remerciements à nos familles pour leur soutien moral et physique durant toutes nos années d'étude ainsi que leur compréhension.

*Nous tiendrons à remercier notre promotrice **Mme. Siham OUAMOUR,** Maitre de Conférences à l'USTHB et notre promoteur **Mr. Halim SAYOUD,** Professeur à l'USTHB, pour nous avoir suivi, orienté et encouragé avec leur modestie et leur serviabilité.*

Une pensée particulière à tous nos enseignants du primaire à l'université pour nous avoir donné un bon niveau d'instruction.

Ainsi, nous présentons toutes nos gratitudes pour tous ceux qui nous ont aidés de près ou de loin par encouragement ou par contribution afin de pouvoir continuer surtout bien nos collègues et nos amis.

Nous remercions également le président et les membres de jury.

LYNDA

et

HADJER

TABLE DES MATIÈRES

Pages

Introduction générale……..…1

Chapitre I : Généralités sur la localisation de source sonore.

I.1- Introduction……...3

Ⅰ.2-C'est quoi un son ?.....3

Ⅰ.3-Physiologie de l'oreille humaine……4

Ⅰ.4-Mécanisme d'audition…….5

Ⅰ.5-La localisation de la source sonore……6

Ⅰ.5.1-Localisation à l'azimut (la direction)……...6

a) Les Différences de Niveau Interauriculaires (DNI)……..6
b) Les Différences de Temps Interauriculaires (DTI)……...8

Ⅰ.5.2-Localisation de la profondeur (l'élévation)……...……..9

Ⅰ.5.3-La perception de la distance de la source…….....10

a) En champ libre……10
b) Dans un espace réverbérant.....10

I.6- Conclusion……....11

Chapitre II : Méthodes de localisation des sources sonores.

II.1- Introduction……12

II.2- Techniques de localisation de la position……...12

II.2.1- Méthode du différentiel d'énergie (EDM)……12

a) Calcul du différentiel d'énergie……12

b) Indication de la position localisée……...13

II.2.2- Méthodes basées sur le calcul du retard……..14

II.2.2.1- Calcul de la distance entre le locuteur et les deux microphones……..14

II.2.2.2- Calcul du retard……..16

a) Méthode de la corrélation…….16

a.1) Définition…….16

a.2) L'autocorrélation « autocorrélation »……..……16

a.3) L'inter-corrélation « cross corrélation »……16

a.4) Estimation du retard par l'inter-corrélation « cross corrélation »……...17

b) Méthode de la corrélation filtrée (FCM)……18

c) Méthode de Generalized cross-correlation phase transform(GCC –PHAT)..19

c.1) Définition……19

c.2) Dérivation de la GCC……20

c.3) Phase Transform (PHAT)……..21

c.4) La GCC – PHAT……..22

II.3- Exemple d'application de localisation ……23

Cas d'un signal sinusoïdal (cas théorique)……..23

II.4- Conclusion……...26

Chapitre III : Base de données expérimentale.

III.1- Introduction……27

III.2-Expériences réalisées……27

III.3- Conception de la base de données DB11……...29

III.3.1- Conditions de réalisation……..29

III.3.2-Scénarios de la base de données BD11……...30

III.3.3- Configurations de la base de données BD11……....31

III.3.4-Discours enregistrés……31

III.4- Dispositifs matériels et logiciels utilisés……32

III.4.1- Schéma synoptique du système……32

III.4.2- Description des différents blocs du système……32

a) Bloc d'acquisition……32

a.1) Microphones……..33

a.2) Préampli audio……...33

b) Bloc de traitement et de commande……35

b.1) Unité de traitement……..35

- Pré-traitement audio avec le logiciel Gold Wave……...35
- Traitement audio avec le logiciel Matlab……....37

b.2) Circuit de puissance……37

c) Bloc d'exécution……37

III.5- Conclusion……..37

Chapitre IV : Expériences et résultats.

IV.1-Introduction……38

IV.2- Procédure des expériences réalisées……38

IV.2.1- Configuration de la base de données BD11……38

IV.2.2- Acquisition des signaux audio……..38

IV.2.3- Pré-traitement audio avec le logiciel Gold Wave.……38

IV.2.4- Traitement audio avec le logiciel Matlab……....40

a) Méthode du différentiel d'énergie EDM……..40

a.1) Algorithme…….....41

a.2) Résultats des expériences réalisées…….....41

b) Méthode de Generalized cross-correlation phase transform(GCC –PHAT)....48

b.1) Algorithme…….50

b.2) Résultats des expériences réalisées……50

IV.3- Statistiques des résultats obtenus……..54

IV.4- Conclusion……57

Conclusion générale……....58

Bibliographie……..59

TABLE DES FIGURES

Figure 1.1: *Ondulation du tympan en fonction du temps pour un son grave et un son aigu*...3

Figure 1.2: *Système auditif humain*……4

Figure 1.3**:***La localisation naturelle de l'espace sonore*……6

Figure 1.4: *La DNI en fonction de l'angle d'azimut pour différentes fréquences*……..8

Figure 1.5**:***Le cône de confusion*……9

Figure 1.6**:***La Fonction de Transfert Anatomique*……10

Figure 1.7:*Estimation de la distance d'une voix parlée ou chuchotée en espace libre*……11

Figure 2.1 : *Schéma représentant un locuteur entre deux microphones*…….....14

Figure 2.2 : *Estimation du retard par la méthode de* corrélation……17

Figure 2.3 : *Estimation du retard par la méthode de corrélation filtrée*…….18

Figure 2.4: *TDOA entre chaque deux* microphone.....19

Figure 2.5 : *Estimation du retard par la méthode de GCC*-PHAT……..22

Figure 2.6 : *Deux sinusoïdes légèrement décalées*……..24

Figure 2.7: *L'allure des différentes techniques pour le calcul du retard*……*25*

Figure 3.1: *Schéma synoptique de la configuration de la meeting-room*……27

Figure 3.2 : *Photos montrant des locuteurs dans des différentes positions*……30

Figure3.3: *Schéma synoptique de la configuration fixe*……*31*

Figure 3.4:*Schéma synoptique du système automatique réalisé*……32

Figure 3.5 ***:*** *Photo des microphones utilisés dans les enregistrements*……33

Figure 3.6: *Schéma électrique du Pré-ampli audio utilisé*……34

Figure 3.7: *Pré-ampli audio utilisé*……34

Figure 3.8: *Photo montrant le pré-traitement en utilisant le Gold Wave*……36

Figure 4.1: *Visualisation des signaux audio*……39

Figure 4.2 :*Résultats de localisation obtenus par la méthode EDM pour scénario 1*…….42

Figure 4.3 :*Résultats de localisation obtenus par la méthode EDM pour scénario 2*…….42

Figure 4.4 :*Résultats de localisation obtenus par la méthode EDM pour scénario 3*…….43

Figure 4.5 :*Résultats de localisation obtenus par la méthode EDM pour scénario 4*……..43

Figure 4.6 :*Résultats de localisation obtenus par la méthode EDM pour scénario 5*……..44

Figure 4.7 :*Résultats de localisation obtenus par la méthode EDM pour scénario 6*……..44

Figure 4.8 *: exemple de localisation d'un locuteur au milieu (scénario 1)*……..45

Figure 4.9*: exemple de localisation d'un locuteur à droite (scénario 1.)*……46

Figure 4.10:*Résultats obtenus par la méthode de la GCC-PHAT pour scénario 1*……..48

Figure 4.11:*Résultats obtenus par la méthode de la GCC-PHAT pour scénario 2*……..48

Figure 4.12:*Résultats obtenus par la méthode de la GCC-PHAT pour scénario 3*……..49

Figure 4.13:*Résultats obtenus par la méthode de la GCC-PHAT pour scénario 4*……..49

Figure 4.14:*Résultats obtenus par la méthode de la GCC-PHAT pour scénario 5*……..50

Figure 4.15:*Résultats obtenus par la méthode de la GCC-PHAT pour scénario 6*……..50

Figure 4.16: *Taux de précision de localisation de la position par scénario*……..52

Figure 4.17: *La moyenne des taux de précision par méthode*……..*53*

LISTE DES TABLEAUX

Tableau 2.1: *Résultats de simulation des différentes méthodes*……25

Tableau 3 .1 : *Récapitulatif des bases de données réalisées*……...28

Tableau 3 .2 : *Résultats de localisation des locuteurs*……28

Tableau 4.1: *Taux de précision de localisation de la position par scénario*……..51

Tableau 4.2: *La moyenne des taux de précision par méthode*……..53

Au cours de ces dernières années, le problème de la localisation d'une source sonore dans l'espace a connu un intérêt croissant par la communauté scientifique qui veut rapprocher de plus en plus de la capacité humaine dans la détection du son (Brutti, 2008).

Plusieurs applications de traitement audio peuvent obtenir des avantages substantiels sur la connaissance de la position de la source qui émet le signal sous processus et dans la plupart de ces applications, on estime cet emplacement. Ces estimations doivent être produites à un rythme élevé et avec une latence très faible, mais ils ont besoin d'être précis.

Pour cette raison, de grands efforts ont été consacrés à enquêter sur ce domaine de recherche et plusieurs approches ont été proposées ce qui a enrichi le domaine de l'intelligence artificielle de point de vue conceptions de matériel ainsi que la mise en œuvre de logiciels temps-réel (Hoang, 2009).

En général, il existe deux catégories d'algorithmes qui ont été explorées pour les systèmes réels: les algorithmes à deux étages et les algorithmes à un étage. L'exécution et le coût de calcul de ces deux catégories varient selon les conditions des environnements, les données étant traitées, les moyens mis en œuvre ces algorithmes et aussi les causeurs eux-mêmes.

- L'algorithme à deux étages se fait en deux étapes :

Dans la première étape, ou l'étape de l'estimation du temps de retard (TDE: Time Delay Estimation), il se produit un temps de retard de l'arrivée (TDOA: Time Difference Of Arrival) du son entre des paires de microphones.
Dans la deuxième étape, les courbes hyperboliques sont générées à partir des estimations disponibles de temporisation et de données de position du microphone, et puis on doit estimer l'emplacement de la source de l'intersection de ces courbes dans un certain sens optimal.

- L'algorithme à un étage ne traite pas d'une manière paire, mais exploite la multitude de microphones, afin de surmonter la limitation indiquée par la prise de décisions rapide et de réverbération (Hoang,2009).

Traditionnellement, les algorithmes de la localisation des sources sonores sont fondés sur une estimation de la différence du temps d'arrivée (TDOA) aux deux paires de microphone à partir du quel on peut déduire des informations sur la position spatiale d'un émetteur source (Brutti, 2008).
Ce type de système basé sur des réseaux de microphones a été utilisé dans de nombreuses applications, telles que la téléconférence, la télésurveillance, la reconnaissance de la parole, la caractérisation du causeur et la capture de la voix dans des environnements réverbérants (Hoang,2009).

C'est pourquoi le but de notre projet est de proposer un système permettant de détecter la position d'un locuteur donné en utilisant son signal de parole enregistré par deux microphones placés en opposition et séparés par une distance fixe .

Généralement, dans les meeting-rooms, les interlocuteurs sont assis autour d'une table à des positions fixes. Ceux-ci se partagent une discussion assez continue et hiérarchique. On utilisera 2 microphones pour la capture du signal de parole. A partir de ce signal, un traitement adéquat permettra de trouver lequel des locuteurs a parlé et où il se trouve.

Plan de travail :

Nous avons organisé notre thèse en quatre chapitres :

- Le premier chapitre porte sur une recherche bibliographique sur le système auditif humain ainsi que le domaine de la localisation spatiale de source sonore.

- Le second chapitre expose certaines techniques de localisation spatiale de la source sonore telles que: la méthode de la corrélation, la méthode de la corrélation filtrée, la méthode à énergie différentielle et GCC-PHAT (Generalised Cross Correlation-PHAse Transform).

- Dans le troisième chapitre, on va discuter la base de données audio réalisée expérimentalement DB11.

- Le dernier chapitre est consacré à l'évaluation de nos algorithmes sur la base de données stéréophonique DB11 et l'interprétation des résultats obtenus.

Enfin, nous terminons par une conclusion générale, des perspectives et des références bibliographiques pouvant éventuellement éclaircir certains points.

Chapitre I

Généralités sur la localisation de source sonore

Chapitre II

Méthodes de localisation des sources sonores

Chapitre III
Base de données expérimentale

Chapitre IV
Expériences et résultats

Ⅰ.1- Introduction

Le système auditif humain constitue un ensemble très performant. En effet, avec seulement deuxrécepteurs (oreilles), le cerveau est capable, et très rapidement, de prendre en compte tous les sons dans notre entourage et de nous indiquer sur leurs positions d'émission.
De nombreuses équipes de recherches dans le monde entier se sont penchées sur l'idée de pouvoir, comme pour l'être humain, localiser la source d'émission d'un son (Margueron, 2000).

Leurs étudesmènent à plusieurs méthodes de localisation spatiale des sources sonores.
Pour y pouvoir étudier ces méthodes, il faut d'abord bien comprendre ce qu'est un son, puis comment on peut localiser leur source.

Ⅰ.2-C'est quoi un son ?

Les sons "purs" sont des oscillations régulières de la pression de l'air (les spécialistes parlent d'oscillations sinusoïdales)(Philipe, 2012). Ces oscillations issues de l'effet des vibrations rapides des corps, se propageant dans les milieux matériels et excitant l'organe de l'ouïe.

Quand un corps sonore a été frappé, ses différentes parties éprouvent aussitôt un mouvement de vibration. L'air qui environne ce corps participe à ce mouvement et forme autour de lui des ondes qui parviennent à l'oreille. L'air est donc le principal véhicule du son, qui se propage avec une vitesse de 340 m/s environ (à la température ordinaire)mais les sons aigus sont des vibrations plus rapides que les graves (leur fréquence est plus élevée).
Les liquides transmettent le son avec plus de rapidité ainsi que dans les solides, la vitesse est encore plus grande.

Le son ne se transmet pas dans le vide, et son intensité augmente ou diminue en même temps que la pression du gaz qui le transmet.
Lorsque les ondes sonores rencontrent des obstacles fixes, elles se réfléchissent. C'est sur cette propriété qu'est fondée la théorie de l'écho (Larousse, 1969).

Voici dans la figure 1.1comment vibre la membrane du tympan dans l'oreille en fonction du temps quand elle reçoit une note de musique(Philipe, 2012) :

-une note à 440 Hz qui est un son grave (un **La**, de la 3ème octave, en bleu).
- un son plus aigu à 1760 Hz(un **La** aussi, mais de la 5ème octave, en rouge).

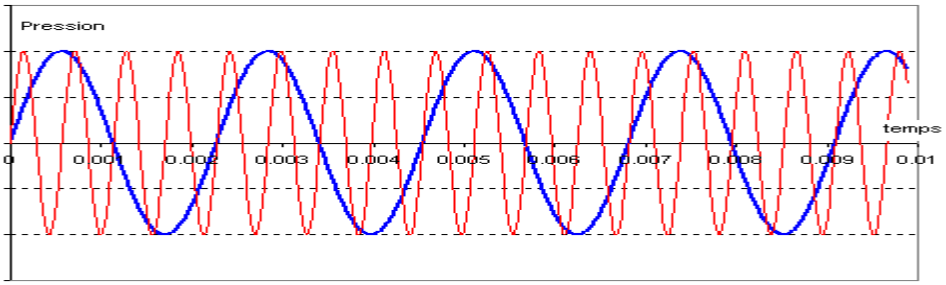


Figure1.1: Ondulation du tympan en fonction du temps pour sons grave et aigu(Philipe, 2012).

Ⅰ.3-Physiologie de l'oreille humaine

L'oreille est un organe jouant le rôle d'un transducteur mécanique quiconvertit l'information de pression en une information électrique(Khennouf, 2012). Elle est constituée chez l'être humain de 3 parties:

a. L'oreille externe, avec le pavillon et le conduit auditif, fermé par le tympan.
b. L'oreille moyenne, cavité ou caisse du tympan, communiquant avec le pharynx par la trompe d'Eustache, et dans laquelle une chaine de trois osselets (marteau, enclume, étrier) transmet les vibrations du tympan à la fenêtre ovale, et puis à l'oreille interne.
c. L'oreille interne, contient l'organe de l'équilibration et l'appareil auditif, formé de limaçon, contenant les cellules auditives ciliées de l'organe de Corti (Larousse,1969).

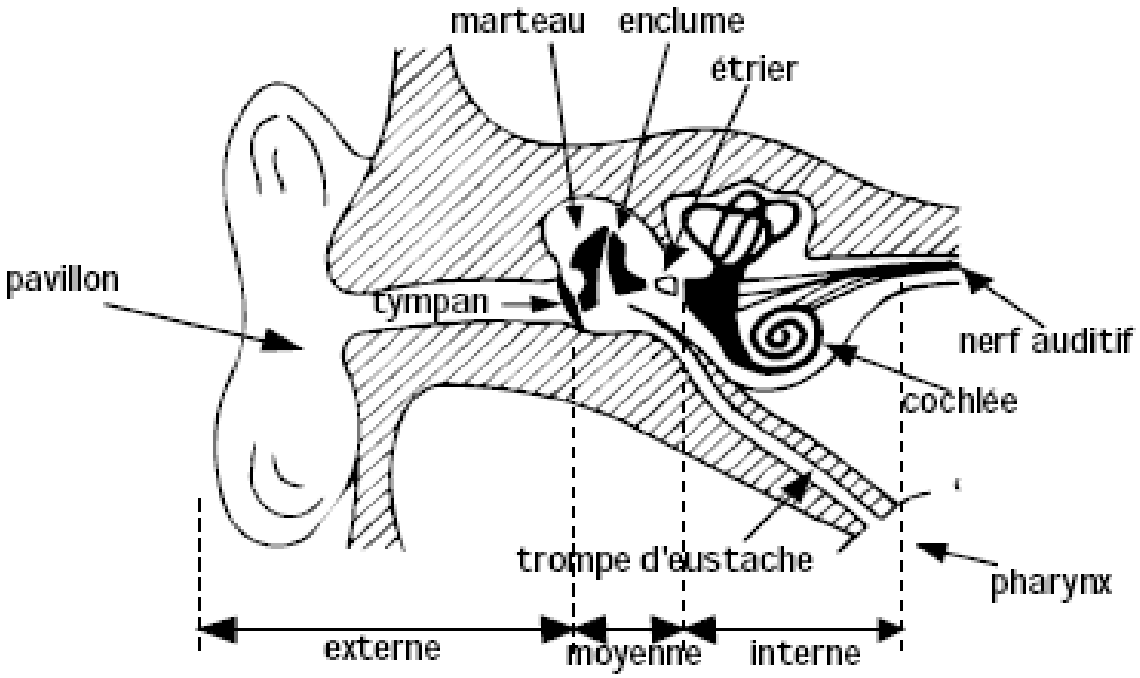


Figure 1.2: Système auditif humain (Ait Saadi , 2005).

Ⅰ.4-Mécanisme d'audition

Les ondes sonores sont recueillies par l'appareil auditif. Ces ondes de pression sont ensuite analysées dans l'oreille interne, puis les résultats sont interprétés par le cerveau.
Le conduit auditif relie le pavillon au tympan : c'est un tube acoustique de section uniforme fermé à une extrémité, son premier mode de résonance est situé à 3 KHz, ce qui accroît la sensibilité du système auditif humain dans cette gamme de fréquences.

Le mécanisme de l'oreille interne (*marteau,étrier, enclume*) permet une adaptation d'impédance entre l'air et le milieu liquide de l'oreille interne. Les vibrations de l'étrier sont transmises au liquide de la *cochlée*. Celle-ci contient la *membrane basilaire* qui transforme les vibrations mécaniques en impulsions nerveuses. La membrane s'élargit et s'épaissit au fur et à mesure que l'on se rapproche de l'apex de la cochlée; elle est le support de l'*organe de Corti* qui est constitué par environ 25000 *cellules ciliées* raccordées au nerf auditif. Chacune de ces cellules présente une réponse en fréquence où la fréquence de résonance dépend de la position occupée par la cellule sur la membrane.

L'oreille humaine est capable d'entendre sur une gamme de fréquence comprise entre les infrasons ($\leq$ 20 Hz) et les ultrasons ($\geq$ 20 kHz).À l'intérieur de cet espace fréquentiel existe un sous-espace délimité par les niveaux d'énergie des sons au-delà de lesquels l'homme ne percevra pas un son d'une fréquence appartenant au spectre de l'audition. Cette limite d'énergie est appelée seuil d'audition et il est variable en fonction de la fréquence. Inversement, il existe une limite d'énergie maximale appelée le seuil de douleur et qui varie aussi en fonction de la fréquence. Cette limite ne doit pas être franchie car la cochlée, et plus particulièrement les cellules ciliées, peuvent être irrémédiablement endommagées(Ait Saadi , 2005).

La présence des deux oreilles permet d'effectuer, au niveau du cerveau, des traitements plus complexes que le simple décodage d'une scène auditive. Le positionnement des oreilles de chaque côté du crâne permet en effet de profiter des capacités de la binauralité. Cette faculté permet de calculer la provenance d'un son en fonction du retard d'arrivée de ce dernier dans une oreille par rapport à l'autre. Il est à noter que cette binauralité permet à l'homme de discerner la position horizontale de l'émetteur d'un son mais pas sa position verticale(Benselama, 2007).

Ⅰ.5-La localisation de la source sonore

La perception des sons dans l'espace est essentiellement reliée à la perception binauriculaire, c'est-à-dire au fait d'avoir deux oreilles. La localisation des sons dans l'espace est complémentaire des autres sens, notamment la vision et la perception de son corps dans l'espace (proprio-perception).
La localisation comprend trois dimensions spatiales: l'azimut ou la direction (gauche-droite), la hauteur ou l'élévation (haut-bas), et la profondeur ou la distance à laquelle elle se trouve (Localisation d'un point d'émission,2012).

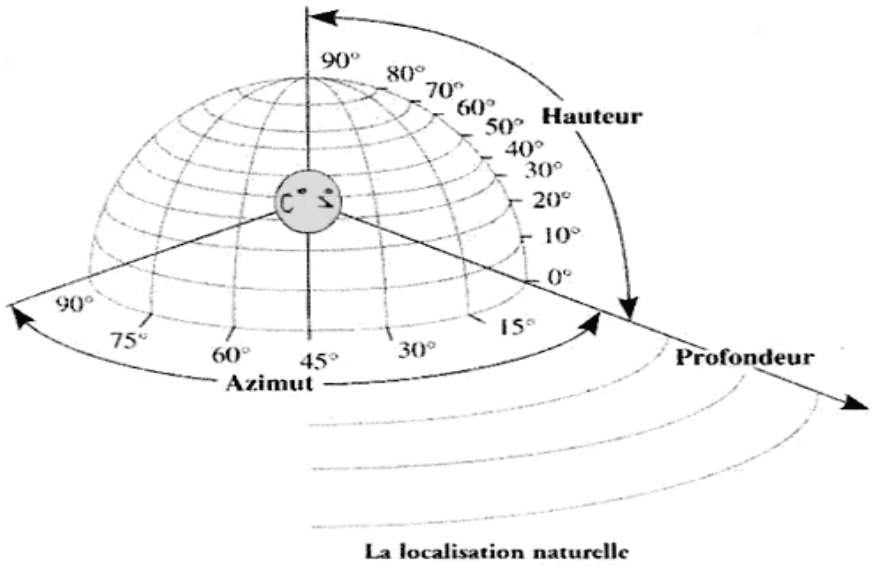


Figure1.3: La localisation naturelle de l'espace sonore (Localisation d'un point d'émission, 2012).

Ⅰ.5.1-Localisation à l'azimut (la direction)

Les indices pertinents sont les différences de niveau interauriculaires(DNI) et les différences de temps interauriculaires(DTI). Ces deux indices sont indépendants (localisation de la source sonore,2012).

a) Les Différences de Niveau Interauriculaires (DNI)

Il y a 120 ans, Lord Rayleigh avait déjà compris l'importance pour la localisation de la différence de niveaux d'intensité arrivant aux deux oreilles.
À cause du phénomène de la diffraction, un son venant de la droite sera plus fort à l'oreille droite qu'à l'oreille gauche, et ce d'autant plus que la fréquence est élevée. En effet, la tête constitue un obstacle à franchir pour les ondes sonores et une "ombre acoustique" apparaît, d'autant plus marquée que la fréquence est élevée(localisation de la source sonore, 2012).

- Calcul de la fréquence critique (correspondant aux dimensions d'une tête) :

$$\text{Diamètre de la tête} = 23\ \text{cm} = 0{,}23 \qquad (1.1)$$

$$f = c / l = 344 / 0{,}23 \approx 1500\ \text{Hz} \qquad (1.2)$$

À 500 Hz, la longueur d'onde d'un son est de 68 cm, ce qui représente plus de 3 à 4 fois le diamètre d'une tête moyenne. À cause de la diffraction, la DNI est donc petite pour des fréquences inférieures à 500 Hz, si la source est au moins à un mètre de distance.

- La DNI est un indice de localisation surtout utilisé pour les hautes fréquences (au-delà de 4000 Hz).

La figure1.4 représente la DNI en fonction de l'angle d'azimut mesuré pour différentes fréquences. On remarque que la différence de niveau s'accentue pour des plus hautes fréquences.

(Localisation de la source sonore, 2012).

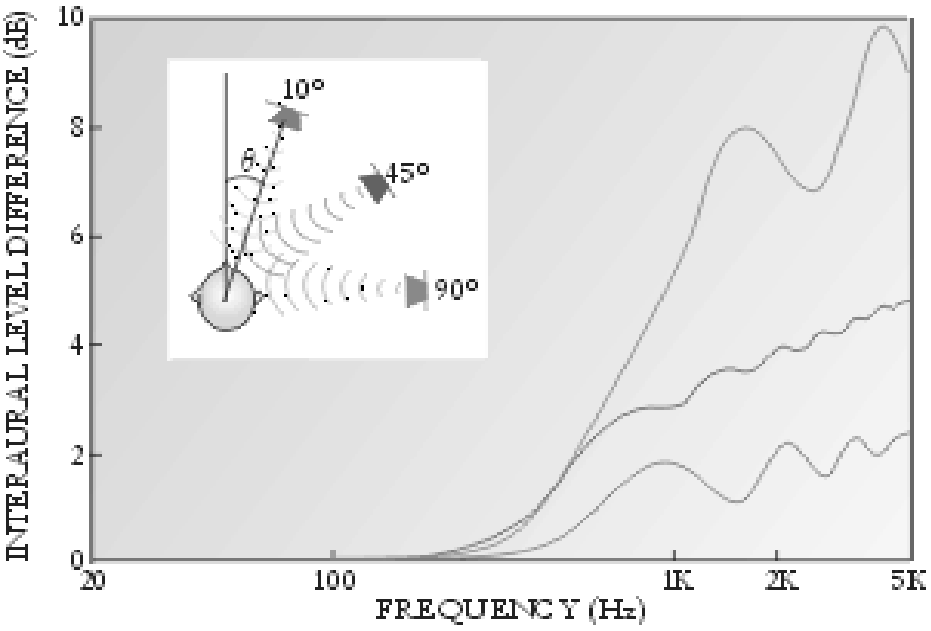


Figure 1.4: La DNI en fonction de l'angle d'azimut pour différentes fréquences(localisation de la source sonore, 2012).

a) Les Différences de Temps Interauriculaires (DTI)

Pour les basses fréquences (grandes longueurs d'ondes →λ > tête), il semble que notre système auditif binauriculaire utilise la différence de phase (Δ φ) ou différence de temps d'arrivée (Δ t) entre les signaux arrivant aux deux oreilles (les signaux étant repérés par les maxima ou les passages à zéro) (localisation de la source sonore, 2012,).

- Une différence de phase Δ φ correspond à une DTI :

$$\Delta t = \Delta\varphi/(2\pi f) \qquad (1.3)$$

- Pour calculer une approximation de la DTI, on considère la demi-circonférence correspondant au parcours de l'onde autour de la tête pour passer d'une oreille à l'autre :

$$d = \pi R \qquad (1.4)$$

OùR est le rayon de la tête.

- La DTI vaut alors :

$$\Delta t = d/V \qquad (1.5)$$

Où v est la vitesse du son dans l'air.

- Calcul de la DTI :

Rayon de la tête = 10 cm = 0,1 m (1.6)

Demi-circonférence = 0,1 π = 0,314 m (1.7)

La plus grande valeur possible (dans le cas où la source est en face d'une des deux oreilles) est :

DTI = Δ t = 0,314 / 340= 924 μs = 0,924 ms (1.8)

Pour les grandes longueurs d'onde, on peut utiliser la formule pour la diffraction par une sphère :

$$\Delta t = \left(\frac{3R}{V}\right) \sin\theta \qquad (1.9)$$

Où θ est l'angle azimutal (gauche-droite).

À 90° (source en face d'une des deux oreilles) :

Δ t = (3 x 0,1 / 340) = 0,882 ms. (1.10)

- Les DTI sont des indices de localisation pour les plus basses fréquences(≤1500 Hz).

(Localisation de la source sonore,2012).

Ⅰ.5.2-Localisation de la profondeur (l'élévation)

Considérant le modèle de la tête sphérique, tous les points compris dans un plan vertical perpendiculaire à l'axe passant par les deux oreilles, correspondent à des signaux arrivant aux deux oreilles avec la même intensité et la même phase. Les indices de DNI et DTI ne permettent donc pas de distinguer le haut, le bas, l'avant et l'arrière, dans le plan vertical.
D'autre part, des sources sonores situées en tout point de la surface d'un cône, appelé cône de confusion(Figure 1.5), causeront des DTI identiques. Sans le support visuel, le son peut être perçu comme provenant du côté opposé sur le cône, par rapport à son point d'origine réel(localisation de la source sonore,2012).

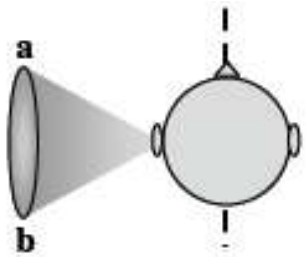


Figure 1.5: Le cône de confusion (localisation de la source sonore,2012).

Malgré tout, la localisation avant/arrière et haut/bas est possible grâce à l'interprétation par le cerveau du filtrage acoustique du son par la Fonction de Transfert Anatomique (FTA)ou en anglais Head-Related Transfer Function (HRTF), correspondant au filtrage par le pavillon, la tête et les épaules. La FTA est illustrée dans la figure 1.6 pour trois positions dans le plan vertical (localisation de la source sonore,2012).

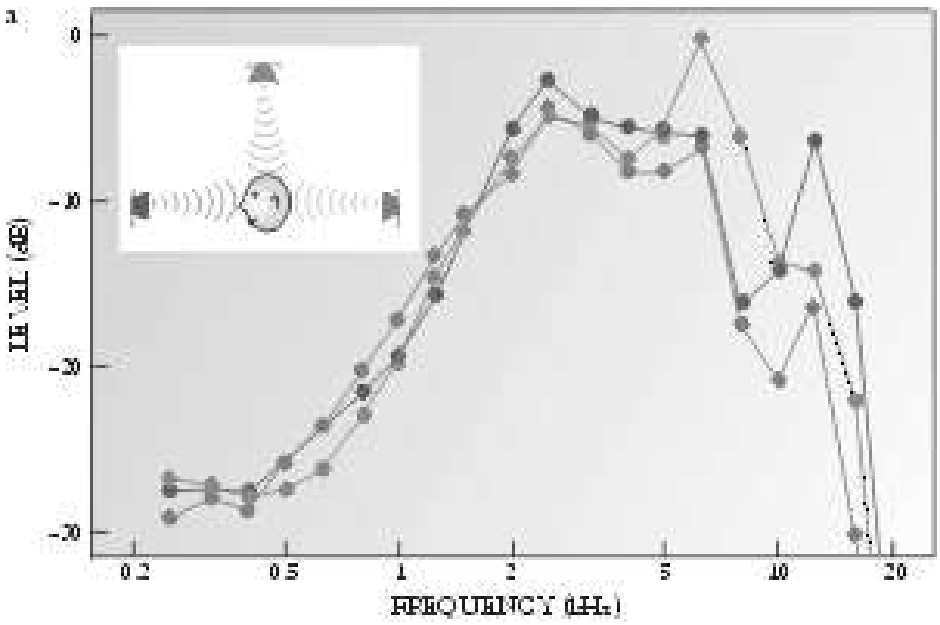


Figure 1.6: La Fonction de Transfert Anatomique (localisation de la source sonore, 2012).

Ⅰ.5.3-La perception de la distance de la source

Lorsqu'on s'éloigne d'une source sonore, l'intensité diminue de 6 dB à chaque doublement de distance.
Diverses expériences ont démontré que des sources sonores familières sont localisées avec plus de précision que des sources inconnues mais ces distances sont toujours sous-estimées. L'estimation de la distance dépend également du niveau d'émission de la source.
En conséquence, si l'on désire recréer artificiellement une impression de profondeur, une atténuation supérieure à 6 dB (jusqu'à 20 dB) est nécessaire pour donner l'illusion d'un doublement de distance(localisation de la source sonore,2012).

a) **En champ libre**

L'analyse spectrale par le système auditif est le principal indice pour la perception de la distance (si la référence connue). En effet, la répartition des graves et des aigus changes avec la distance. Cette faculté est plus précise si la source est en mouvement (localisation de la source sonore, 2012).

b) **Dans un espace réverbérant**

Dans une salle, lorsqu'un auditeur s'éloigne d'une source sonore, le rapport du son direct au son réverbéré diminue.
Le son réverbéré est constitué de multiples réflexions ; il devient rapidement homogène en tous points de la salle. En fait, seule l'énergie du son direct décroît de 6 dB à chaque doublement de distance, d'où une même décroissance du rapport champ direct / champ réverbéré(Localisation d'un point d'émission, 2012).

Notre perception de la distance peut aussi être influencée par notre expérience avec un type de source particulier tel que la voix. Dans une expérience, on avait présenté des voix criées, parlées, et chuchotées à des pressions sonores identiques. Les voix criées étaient systématiquement évaluées comme étant plus distantes, suivies des voix parlées, puis des voix chuchotées(localisation de la source sonore,2012).

La densité spectrale d'un signal sonore varie lors de sa propagation en fonction d'une absorption inégale des fréquences graves et aiguës.
Des expériences ont montré notamment qu'en champ libre, des sons dont le contenu fréquentiel est inférieur à 2 kHz semblent plus éloignés que les sons de fréquence supérieure
La figure 1.7ci-dessous présente les résultats de l'estimation de la distance d'une voix parlée masculine chuchotée, à spectre essentiellement aigu, comparée à celle d'une voix parlée normalement à large spectre. On constate que la perception de la distance d'une voix chuchotée est sous-estimée.
On en déduit que la sensation d'une source sonore proche ou lointaine pourra être créée artificiellement par un filtrage approprié(Localisation d'un point d'émission,2012).

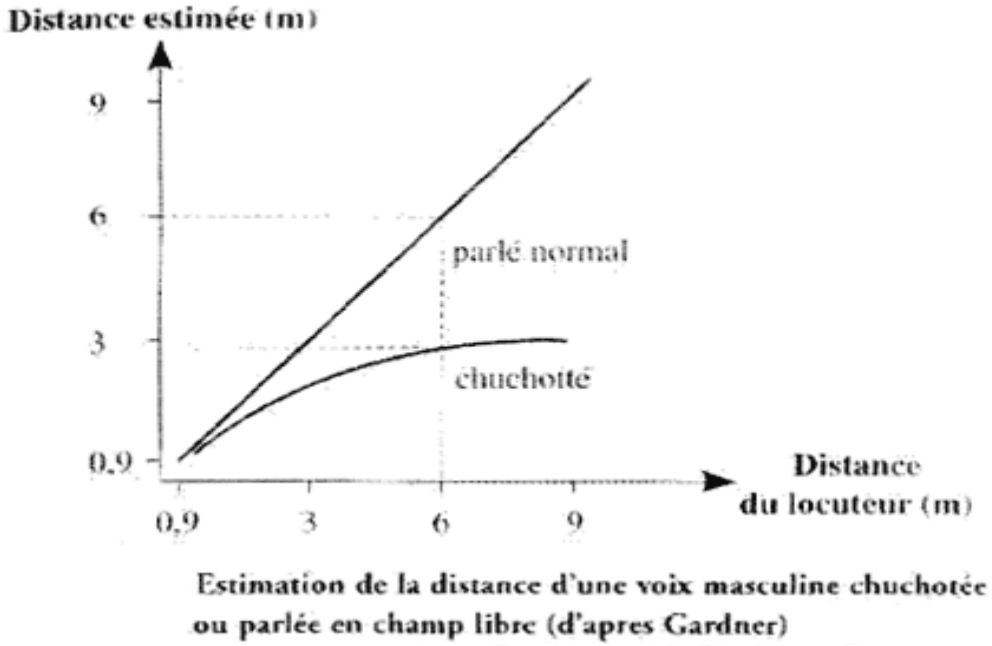


Figure 1.7: Estimation de la distance d'une voix parlée ou chuchotée en espace libre

Ⅰ.6- Conclusion

Dans ce premier chapitre, on a juste jeté un coup d'œil sur le système auditif humain et le mécanisme d'audition ainsi on a rappelé de quelques théories sur lesquelles est basée la localisation de sources sonores.

II.1- Introduction

Dans ce chapitre, on va proposer quelques méthodes de localisation de sources sonores telles que : la corrélation et en particulier la corrélation filtrée, la méthode du différentiel d'énergie et enfin la GCC-PHAT.

L'objectif de toutes ces méthodes est d'aboutir aux mêmes performances que le système auditif humain en termes d'extraction, reconnaissance et localisation.
La comparaison de ces différentes techniques de localisation proposées, nous permet d'établir les avantages et les inconvénients d'une approche donnée dans une mise en œuvre en temps réel **(**Brutti, 2008**).**

II.2-Techniques de localisation de la position

La localisation peut se faire soit par le calcul de la différence d'énergie des signaux arrivant aux deux microphones, ou par le calcul du retard entre les deux signaux de parole collectés par les deux microphones puis déduire la distance entre le locuteur et les deux microphones.

II.2.1- Méthode du différentiel d'énergie (EDM)

a) Calcul du différentiel d'énergie

- Soit un signal s(n) définie sur un intervalle de taille **N** c-a-d [1 N], l'énergie du signal s(n) dans le cas discret peut s'écrire comme suit :

$$E_s=\sum_1^N |S_i|^2 \qquad (2.1)$$

- Dans notre application, on a utilisé l'énergie modifiée qui est calculée comme étant la somme des valeurs absolues des valeurs du signal le long d'une fenêtre d'analyse de taille **N**. Cette énergie est appelée : Energie d'ordre 1 (Khennouf, 2010).

- Les deux énergies (énergie du signal perçu par le microphone droit et l'énergie du signal perçu par le microphone gauche) sont données par les relations suivantes:

$$E_{Sd}=\sum_1^N |X_i| \qquad (2.2.a)$$

$$E_{Sg}=\sum_1^N |Y_i| \qquad (2.2.b)$$

- L'énergie des deux signaux en dB est donnée par :

$$E_{Sd}(dB) = 10*\log_{10}(E_{Sd}) \quad (2.3.a)$$

$$E_{Sg}(dB) = 10*\log_{10}(E_{Sg}) \quad (2.3.b)$$

- Le différentiel de l'énergie logarithmique DE est :

$$DE = E_{Sd}(dB) - E_{Sg}(dB) \quad (2.4)$$

b) Indication de la position localisée

Pour la méthode **EDM,** le principe est simple afin d'estimer la position relative du locuteur par rapport à la position des microphones.

Après avoir calculé la différence d'énergie logarithmique DE, on compare cette différence à un seuil de détection pour localiser le locuteur qui parle.

Ceci se fait suivant l'algorithme ci-dessous:

```
% indication sur la position selon la valeur de DE

Si DE > seuil_max
pos = 'droite'

Sinon Si DE < seuil_min
pos = 'gauche'

Sinon
pos = 'milieu'
Fin Si;
```

Avec les valeurs de **$seuil_{max}$** et de **$seuil_{min}$** sont déterminées expérimentalement.

II.2.2- Méthodes basées sur le calcul du retard

Comme deuxième technique de localisation, on propose le calcul de la distance entre la source sonore (le locuteur qui parle dans notre étude) et les deux microphones placés l'un à droite du locuteur et l'autre à gauche.
Le calcul de cette distance est basé essentiellement sur le calcul du retard.

II.2.2.1. Calcul de la distance entre le locuteur et les deux microphones

Pour cela, on considère un locuteur entre deux microphones espacés tel qu'il est illustré dans la figure 2.1, dont les différents paramètres 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.
- $\boldsymbol{L}$ **:** La distance entre les deux microphones.
- $\boldsymbol{V_S}$**:** La vitesse du son qui est environ 340 m/s dans l'air et à une température ordinaire de 15°C

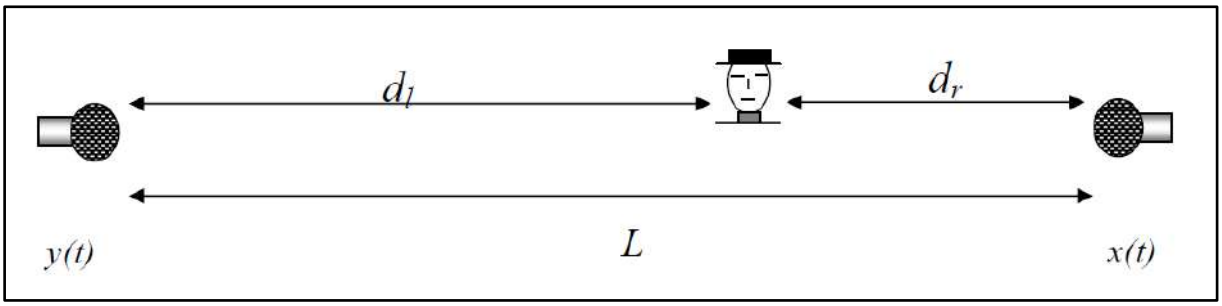


Figure 2.1: Schéma représentant un locuteur entre deux microphones (Khennouf, 2010).

Par exemple, on calcule la distance d_r entre le locuteur et le microphone droit :

$$\mathrm{d_r} + \mathrm{d_1} = \mathrm{L} \quad (2.5)$$

$$\mathrm{delay} = (\mathrm{d_r} - \mathrm{d_1})/\mathrm{V_S} \quad (2.6)$$

$$\mathrm{d_r} - \mathrm{d_1} = \mathrm{delay} * \mathrm{V_S} \quad (2.7)$$

$$\mathrm{d_r} = (\mathrm{L} + \mathrm{delay} * \mathrm{V_S})/2 \quad (2.8)$$

Selon cette distance calculée, on peut détecter la position du locuteur qui parle.

Ceci se fait suivant l'algorithme ci-dessous:

```
%indication sur la position

Si d<seuil_min
pos='droite'

Sinon Si d>seuil_max
pos='gauche'

Sinon
pos='milieu'
Fin Si;
```

Avec les valeurs de **$seuil_{max}$** et de **$seuil_{min}$** sont déterminées expérimentalement.

II.2.2.2.Calcul du retard

a) Méthode de la corrélation(CM)

a.1) Définition

La corrélation est un concept issu de la biologie. C'est par le biais des travaux de Francis Galton (Corrélation, wikipédéa, 2012) que la corrélation devient un concept statistique.

En probabilités et en statistiques, étudier la **corrélation** entre deux ou plusieurs variables aléatoires c'est étudié l'intensité de la liaison qui peut exister entre ces variables (Corrélation, wikipédéa, 2012)

La fonction de corrélation temporelle est définie par la mesure de similitude entre deux signaux, dont l'un est décalé dans le temps par rapport à l'autre. On distingue deux types de corrélation : l'autocorrélation et l'intercorrélation (Khennouf, 2010).

a.2) L'autocorrélation

L'autocorrélation est un outil mathématique souvent utilisé en traitement du signal, consiste à comparer un signal par lui-même décalé avec un certain retard **τ**. L'autocorrélation

permet de détecter des régularités, des profils répétés dans un signal comme un signal périodique perturbé par beaucoup de bruit (autocorrelation, wikipédéa, 2012).

La fonction d'autocorrélation est définie par la relation suivante :

$$cor_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t+\tau)dt \qquad (2.9)$$

a.3) L'inter-corrélation « cross corrélation »

De sa part, l'inter-corrélation est un outil mathématique qui consiste à comparer deux signaux différents dont l'un est décalé avec un certain retard $\boldsymbol{\tau}$ par rapport à l'autre.

La fonction d'intercorrélation est définie par la relation suivante :

$$cor_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)y(t+\tau)dt \qquad (2.10)$$

a.4) Estimation du retard par l'inter-corrélation « cross corrélation »

L'inter-corrélation était la première approche pour estimer le retard entre les signaux reçus par les deux microphones.
Dans le cas discret, la fonction de cross corrélation est donnée par la relation suivante :

$$cor_{dg}(\tau) = \sum_{1}^{N} S_d(n) S_g(n+k) \qquad (2.11)$$

Avec : $S_d(n)$: Signal enregistré par le microphone droit.

$S_g(n)$: Signal enregistré par le microphone gauche.

$$k = \tau . F_e \qquad (2.12)$$

Où F_e : est la fréquence d'échantillonnage.

La figure 2.2 montre la structure de l'inter-corrélation.

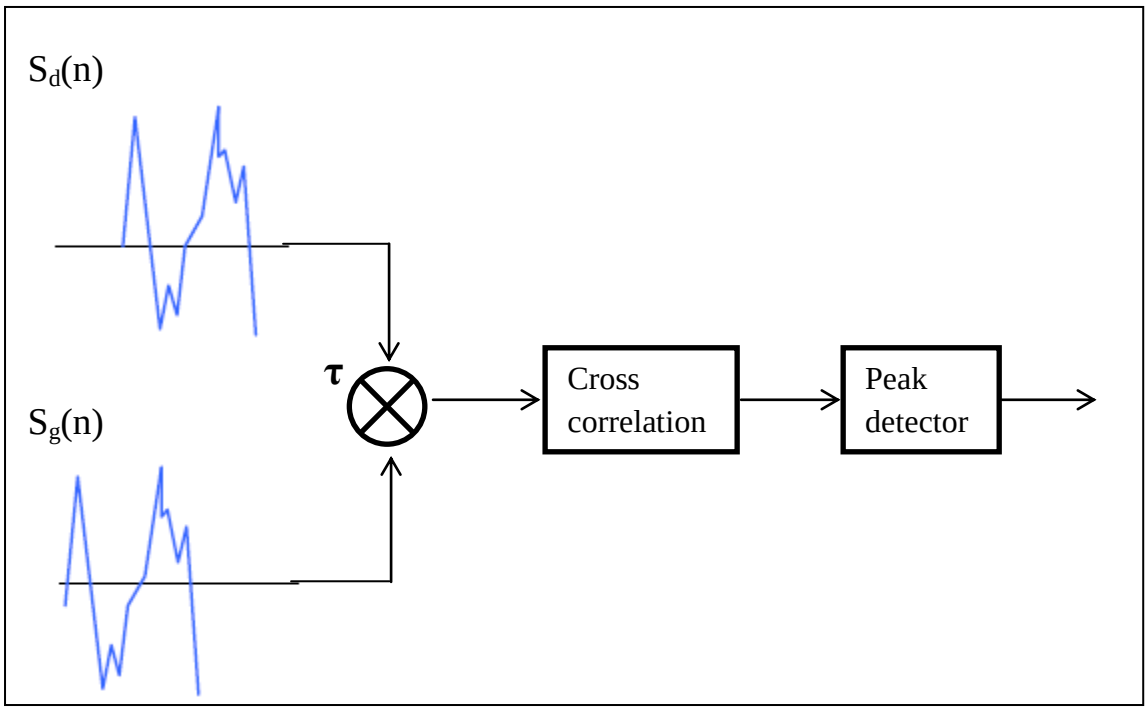


Figure 2.2 : Estimation du retard par la méthode de corrélation.

- L'emplacement (indice) du pic maximal dans la sortie représente le retard estimé :

$$\tau = argmax(cor_{dg}(k)) \quad (2.13)$$

b) Méthode de la corrélation filtrée (FCM)

La méthode de corrélation filtrée **FCM** se fait par deux étapes :

- On calcul d'abord la corrélation discrètetelle quelle est calculée par la relation (2.11).
- Ensuite, on applique au signal $cor_{dg}(\tau)$un filtre passe-bande elliptique, dont la syntaxe est comme suit:

$$f = ellip(N, R_p, R_s, F_p) \quad (2.14)$$

Avec :

N : ordre du filtre.

Rp : atténuation maximale (en dB) dans la bande passante.

Rs : atténuation minimale (en dB) dans la bande coupée.

Fp =2*fc où fc : est la fréquence de coupure.

La figure 2.3 montre la structure de la corrélation filtrée FCM :

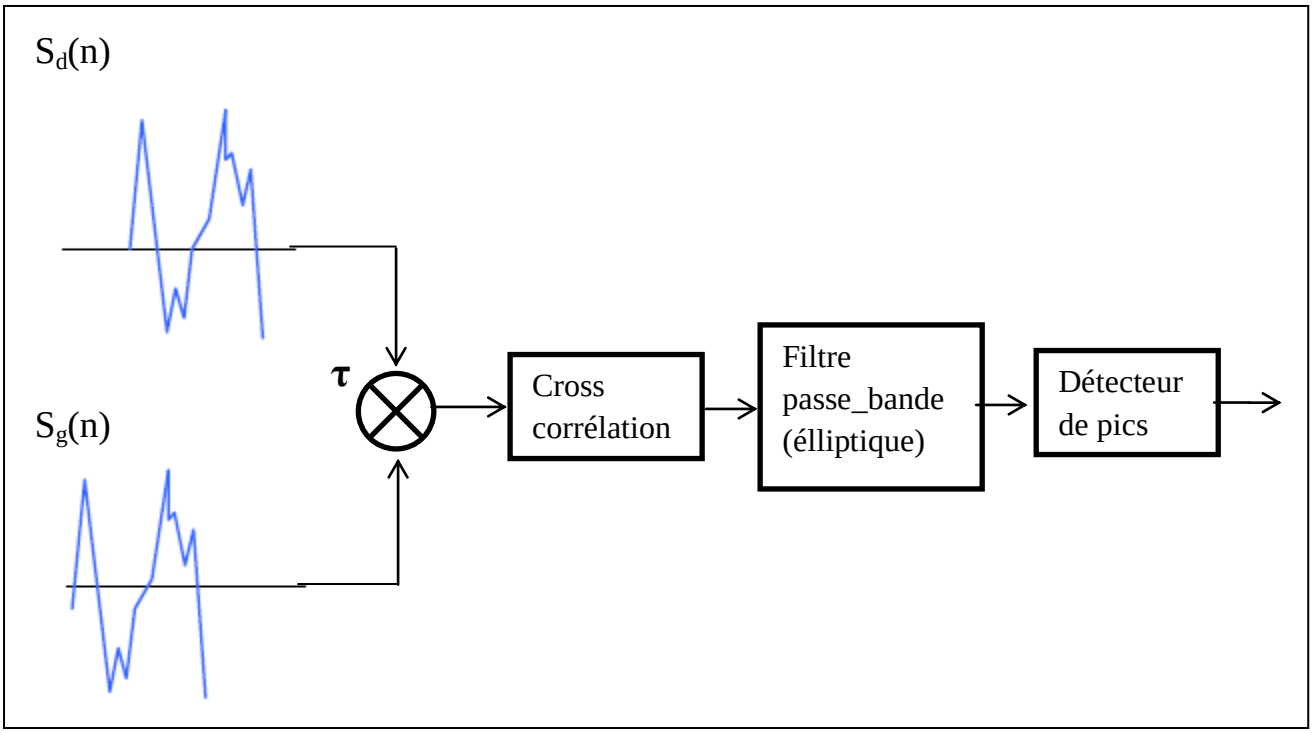


Figure 2.3 : Estimation du retard par la méthode de corrélation filtrée.

- L'emplacement (indice) du pic maximal dans la sortie représente le retard estimé :

$$\tau = argmax(fcor_{dg}(k)) \quad (2.15)$$

c) Méthode de Generalized Cross-Correlation PHase Transform (GCC –PHAT)

c.1) Définition

La GCC-PHAT est parmi les méthodes les plus utilisées pour déterminer la différence de temps d'arrivée (TDOA) «Time Difference Of Arrival» entre deux microphones. Donc, à l'aide de plusieurs valeurs de TDOA, on peut estimer l'emplacement de la source.

Par exemple : on considère un système composé d'une source et 4 microphones.

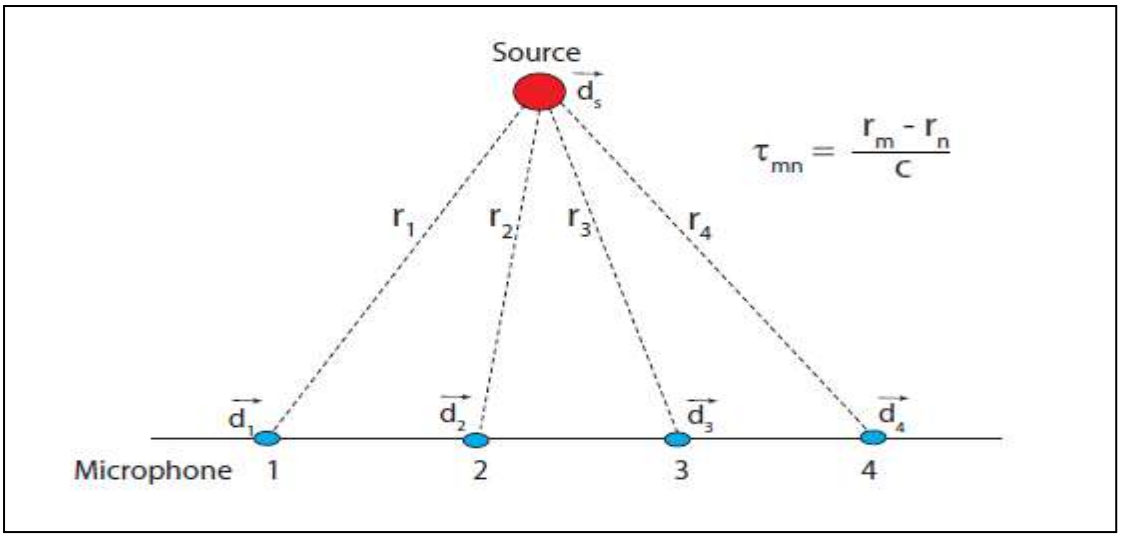


Figure 2.4: TDOA entre chaque deux microphone.

Si la distance entre la source et le microphone est r_m(m=1, 2, 3 ,4) ; le temps de parcours du signal sonore de la source au microphone est :

$$\tau_m = \frac{r_m}{V_S} \qquad (2.16)$$

Où$V_S$est la vitesse du son qui est environ 340 m/s dans l'air et à une température ordinaire de 15°C

Donc, le temps de retard TDOA entre deux microphones m et n peut être défini comme :

$$\tau_{mn} = \tau_m - \tau_n = \frac{r_m - r_n}{V_S} \qquad (2.17)$$

De cette relation entre la TDOA et les distances entre la source et les microphones ,r_m, on peut estimer l'emplacement source**(**Hoang, 2009).

c.2) Dérivation de la GCC

Considérons deux signaux $S_d(t)$ et $S_g(t)$ (signaux reçus respectivement par le microphone droit et le microphone gauche), dont $S_g(t)$ est retardé d'une valeur $\boldsymbol{\tau}$ par rapport à $S_d(t)$.
Donc, on peut écrire :

$$S_g(t) = S_d(t + \tau) \qquad (2.18)$$

Calculons maintenant la différence du temps d'arrivée$\boldsymbol{\tau}$, entre les deux microphones.

- L'inter-corrélation des deux signaux sera montrée un pic correspondant à$\boldsymbol{\tau}$.

Cette inter-corrélation est définie comme suit :

$$cor_{dg}(\tau) = \int_{-\infty}^{+\infty} S_d(t)\ S_g(t)dt = \int_{-\infty}^{+\infty} S_d(t)\ S_d(t+\tau)dt \qquad (2.19)$$

- On prend la transformée de Fourier de l'inter-corrélation $COR_{dg}(\omega)$

$$COR_{dg}(\omega) = \int_{-\infty}^{+\infty} cor_{dg}(t)e^{j\omega\tau}\ \ d\tau \qquad (2.20)$$

- Appliquons la propriété de la convolution de transformée de Fourier sur (2.18) lors de sa substitution en (2.19), on aura :

$$COR_{dg}(\omega) = S_d(\omega).S_g^*(\omega) \qquad (2.21)$$

Avec :

$S_d(\omega)$et$S_g(\omega)$: sont respectivement les transformée de Fourier des signaux $S_d(t)$et $S_g(t)$. $S_g^*(\omega)$est le conjugué de $S_g(\omega)$.

- La transformée de Fourier inverse de(2.20) nous donne la fonction de l'inter-corrélation

$$cor_{dg}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_d(\omega).S_g^*(\omega)e^{j\omega\tau}\ \ d\omega \qquad (2.22)$$

- L'inter-corrélation généralisée GCC de $S_d(t)$et $S_g(t)$ est l'inter-corrélation de leurs deux versions filtrées.
  En désignant les transformées de Fourier de ces deux filtres $F_d(\omega)$ et$F_g(\omega)$, on aura la GCC,$R_{dg}(\tau)$qui est défini comme,

$$R_{dg}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}(F_d(\omega).S_d(\omega)).(F_g(\omega).S_g(\omega))^* e^{j\omega\tau}\ \ d\omega \qquad (2.23)$$

On définit une fonction de pondération combinée $\Psi_{dg}(\omega)$comme,

$$\Psi_{dg}(\omega) = F_d(\omega).F_g^*(\omega) \qquad (2.24)$$

En remplaçant (2.23) dans (2.24), la GCC devient :

$$R_{dg}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Psi_{dg}(\omega).S_d(\omega).S_g^*(\omega)e^{j\omega\tau}\ d\omega \qquad (2.25)$$

(Hoang, 2009)

c.3) Phase Transform (PHAT)

Le TDOA entre les deux microphones est le laps de temps $\boldsymbol{\tau}$ qui maximise la GCC, $\boldsymbol{R_{dg}(\tau)}$.
En réalité,$\boldsymbol{R_{dg}(\tau)}$a de nombreux maxima locaux rendant ainsi plus difficile à détecter la vraie valeurs maximale.
Le choix des fonctions de pondération, $\Psi_{dg}(\omega)$agit sur la performance de la GCC.

Il a été montré que la fonction de pondération de phase transform(PHAT) étudiée dans des conditions réverbérant-libre est robuste dans des environnements réalistes. Cette fonction PHAT qui donne le maximum optimal est définie comme suit,

$$\Psi_{dg}(\omega) = \frac{1}{|S_d(\omega).S_g^*(\omega)|} \qquad (2.26)$$

(Hoang, 2009)

c.4) La GCC - PHAT

En remplaçant la fonction de pondération PHAT de l'équation (2.25) dans l'expression de la GCC de l'équation(2.24), on obtient la GCC-PHAT qui est définie par :

$$R_{dg}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{1}{|S_d(\omega).S_g^*(\omega)|} S_d(\omega).S_g^*(\omega)e^{j\omega\tau}\ d\omega \qquad (2.27)$$

Ou bien, en d'autre terme: [brut]

$$R_{dg}(\tau) = ifft\left(\frac{S_d(\omega).S_g^*(\omega)}{|S_d(\omega)||S_g(\omega)|}\right) \quad (2.28)$$

(Hoang, 2009)

La figure 2.5 montre la structure de la GCC-PHAT :

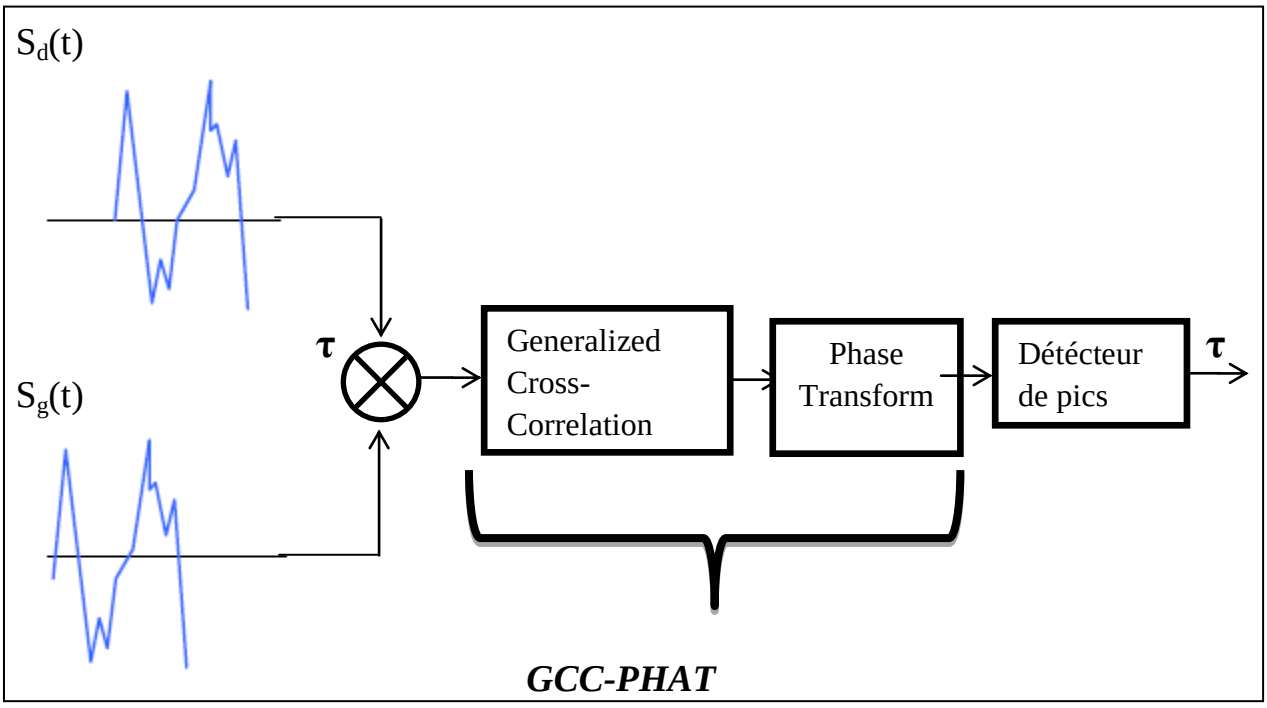


Figure 2.5 : Estimation du retard par la méthode de GCC-PHAT.

- L'emplacement (indice) du pic maximal dans la sortie représente le retard estimé

$$\tau = argmax(R_{dg}(\tau)) \quad (2.29)$$

II.3- Exemple d'application de localisation

- **Cas d'un signal sinusoïdal (cas théorique)**

Avant d'appliquer les différentes méthodes sur notre signal de parole (Voir chapitre 4) et afin d'étudier leurs performances, on a proposé de les tester d'abord sur deux sinusoïdes légèrement décalées dans le temps.

a) Méthode du différentiel d'énergie EDM

On génère de la même source deux sinusoïdes de même période et de même amplitude.

Exemple :

Le capteur de droite reçoit le signal :

a=sin(2*pi*t*fo).*exp(-0.0005*t)+bruit;

Le capteur du gauche reçoit le même signal mais qui a subi une atténuation

b=0.4*(sin(2*pi*t*fo).*exp(-0.0005*t)+bruit);

Leurs énergies seront calculées comme suit :

$E1=\sum_1^N |a(t)|^2$

$E2=\sum_1^N |b(t)|^2$

- L'énergie des deux signaux en dB est donnée par :

$$E1(dB) = 10*\log_{10}(E1)$$

$$E2(dB) = 10*\log_{10}(E2)$$

- Le différentiel de l'énergie logarithmique DE est :

$$DE = E1(dB) - E2(dB)$$

Selon le signe de DE qu'on détermine la position de la source.

- Après simulation, on a trouvé : **DE= +8,399> 0(** E1(dB) >E2(dB)**)**,

Donc la source est plus proche du capteur qui se trouve au coté droit.

b) Méthodes basées sur le calcul du retard (CM, FCM, GCC-PHAT)

On génère de la même source deux sinusoïdes de même période et de même amplitude mais légèrement décalées dans le temps (décalage =23 échantillons).

```
a=sin(2*pi*t*fo).*exp(-0.0005*t)+bruit;
b=sin(2*pi*(t-decalage)*fo).*exp(-0.0005*(t-decalage))+bruit_decale;
```

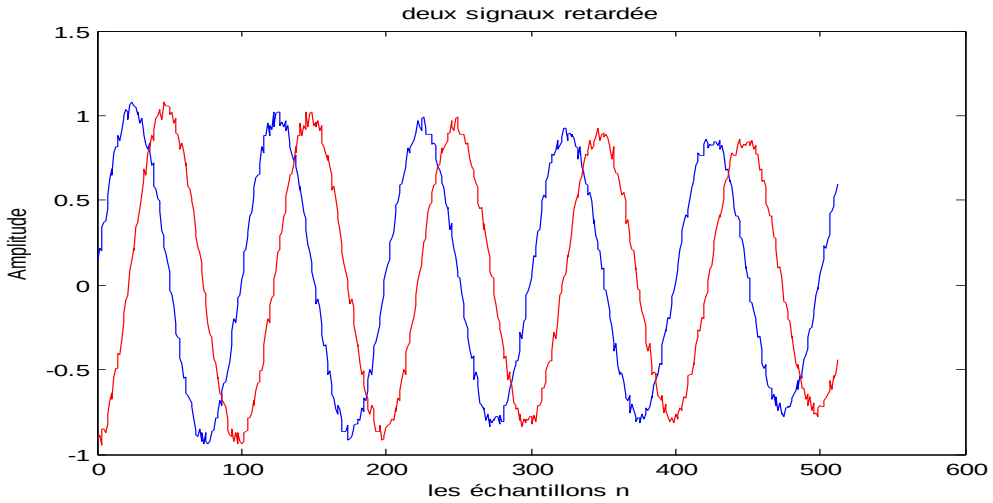


Figure 2.6 : Deux sinusoïdes légèrement décalées.

b.1) Calcul du retard

- Pour la méthode de la corrélation **CM** : on a utilisé l'instruction MatLab c=xcorr(a,b)
- Pour la méthode de la corrélation filtrée **FCM** : on a filtré le résultat de la méthode précédente par un filtre passe-bande de type elliptique [f=ellip (N, Rp, Rs, Fp)]

Avec :

N : ordre du filtre. (On prend N=4)

Rp : atténuation maximale (en dB) dans la bande passante. (On prend Rp=0.5 dB)

Rs : atténuation minimale (en dB) dans la bande coupée. (On prend Rs=60 dB)

Fp =2*fc avec fc=0.25Hz (0<fc<fe/2 et fe=1Hz)

Donc :

f=ellip(4,0.5,60,0.5);

- Pour la méthode de la **GCC-PHAT** : on a utilisé la formule(2.28).

La figure 2.7 illustrent les différentes méthodes de localisation :

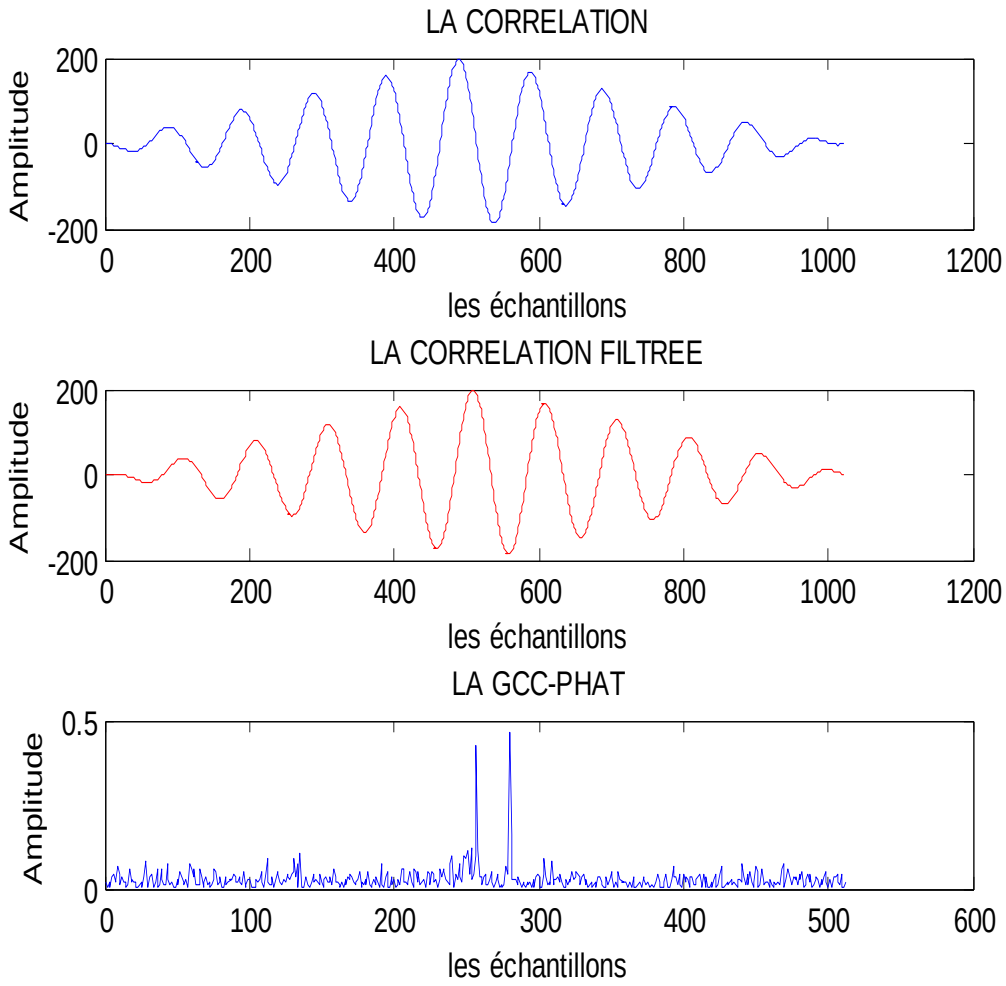


Fig 2.7: L'allure des différentes techniques pour le calcul du retard.

- L'emplacement du pic maximal dans les différents graphes représente le retard estimé.

Après simulation par MatLab, on a obtenues les résultats illustrés dans le tableau suivant :

Tableau 2.1: Résultats de simulation des différentes méthodes.

La méthode	**Le retardτ (échantillon)**
CM	23
FCM	23
GCC-PHAT	23

Commentaire

Théoriquement, toutes les méthodes présentent une bonne performance dans le calcul du retard entre les deux signaux.

b.2) Localisation de la position

Comme on a démontré dans le paragraphe (**II.2.2.1.)** de ce chapitre, calculer la distance revient à calculer le retard. Cette distance va nous indiquer sur la position de la source selon la relation (2.8).

Avec :

$$\tau \text{ [Second]} = \tau \text{ [échantillon]} / Fe \qquad (2.30)$$

Où Fe est la fréquence d'échantillonnage.

II.4- Conclusion

Dans ce chapitre, on a proposé quelques méthodes utilisées pour la localisation des locuteurs telles que la corrélation filtrée, le différentiel d'énergie et la GCC-PHAT.
Afin d'étudier leurs performances, on les a appliqué d'abord sur un signal sinusoïdal et après avoir les testé, on a trouvé que pratiquement toutes les techniques ont donné des très bons résultats de localisation.

III.1- Introduction

Pour les tests d'évaluation des performances de notre système, la conception d'une base de données constituée de plusieurs scénarios avec différentes configurations était nécessaire. Les enregistrements audio ont été faits à l'aide d'un systèmed'acquisition réalisé.
Dans ce chapitre, on va présenter toutes les étapes, le matériel et les logiciels utilisée pour réaliser et traiter cette base de données.

III.1- Expériences réalisées

L'objectif de notre travail est la conceptiond'un système automatique intelligent afin d'orienter une caméra mobile vers des cibles sonores (des locuteurs qui parlent dans notre application).
Ces locuteurs sont assis derrière une table comme illustré dans la figure 3.1suivante :

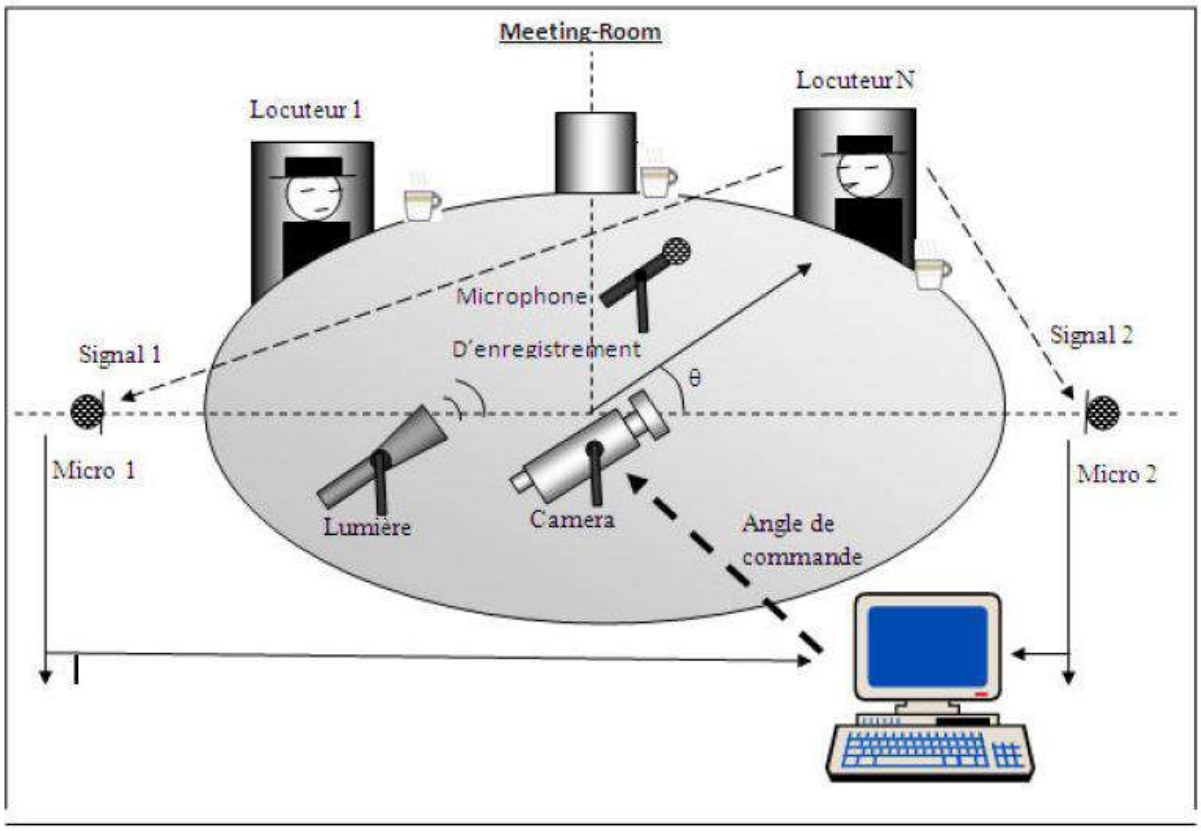


Figure 3.1:Schéma synoptique de la configuration de la meeting-room.

Plusieurs expériencesont été effectuéesenchangeant à chaque fois différents paramètres dans le but de réaliser une base de données signifiante avec de bonnes performances.

Ci-dessous un tableau récapitulatif des différentes bases de données réalisées :

Tableau 3.1 : Récapitulatif des bases de données réalisées.

Bases de données	Lieu des enregistrements	Distance entre microphones (m)	Locuteurs masculins	Locuteurs féminins	Parole préenregistrée	Nombre de scénarios	Durée moyenne des scénarios (s)	Interlocuteurs par scénario	Positions fixes
BD 8	-laboratoire non vide. -Non sourd. -De: 5m x 3m	**4,20**	**3**	**2**	**0**	**27**	**7**	**1 ou 2**	-Gauche -Milieu_G -Milieu_D -Droite
BD 9	-chambre non vide. -mur couverts. -De: 3,8m x 3,4m	**2**	**1**	**1**	**1**	**15**	**22**	**1 ou 2**	-Gauche -Milieu -Droite
BD 10	-chambre non vide. -Non sourde. -De: 3,8m x 3,4m	**1**	**1**	**1**	**1**	**19**	**23**	**1 ou 2**	-Gauche -Milieu -Droite
BD 11	-chambre non vide. -Non sourde. -De: 3,8m x 3,4m	**1**	**2**	**2**	**1**	**6**	**94**	**1, 2 ou 3**	-Gauche -Milieu -Droite

- **Observation des résultats**

Les résultats de localisation des locuteurs sur les différentes bases de données réalisées sont illustres dans le tableau 3.2 suivant:

Tableau 3.2 : Résultats de localisation des locuteurs

Base de données	Résultats de localisation des locuteurs
BD 8	Echec total
BD 9	Faibles performances
BD 10	Bonnes performances, base de données non signifiante
BD 11	Bonnes performances, base de données signifiante

En examinant ces résultats on peut tirer les conclusions suivantes :

- Avec les deux bases de données BD8 et BD9, nous avons obtenus des très mauvais résultats qui sont dû probablement à l'effet d'écho.

- Dans la base de données BD10, le locuteur est localisé correctement, mais le nombre réduit des locuteurs (deux locuteurs) a fait que la base de données ne soit signifiante.

- Nous avons gardés la dernière base de données (BD11), car la localisation du locuteur est faite correctement et elle contient un nombre signifiant de locuteurs (3 locuteurs) et une variété de scénarios (6 scénarios).

III.3- Conception de la base de données BD11

III.3.1- Conditions de réalisation

D'après le tableau 3.1,cette expérience porte les données suivantes :

- La distance entre les deux microphones est : 1m.
- Lieu des enregistrements : chambre non vide, non sourde de 3,8m x 3,4m.
- Nombre de locuteurs masculins : 2.
- Nombre de locuteurs féminins : 2.
- Parole enregistrée : 1.
- Nombre de scenarios est : 6 scenarios.
- Nombre d'interlocuteurs par scenario : 1,2 ou 3 personnes différentes.
- Durée moyenne des scénarios est : 3mn 34s.
- Positions fixes : droite, milieu et gauche.

III.3.2- Scénarios de la base de données BD11

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 un locuteur masculin et un locuteur féminin.
- La troisième catégorie contient un discours entre plusieurs interlocuteurs (masculins et féminins) qui prennent la parole alternativement dans une manière naturelle.

Voici dans la figure 3.2 quelques images montrant certains scénarios de notre BD11 :



a)locuteur à gauche b) locuteur au milieu c) locuteur à droite



d) un locuteur à gauche et un à droite e) un locuteur à gauche et un au milieu

Figure 3.2: Photos montrant des locuteurs dans des différentes positions.

III.3.3- Configurations de la base de données BD11

Deux configurations possibles peuvent être proposéespour les différents scénarios: Une configuration fixe et une configuration mobile.
Dans la configuration fixe le locuteur est assis au tour d'une table dans une des trois positions fixes : gauche, milieu ou droite. Cependant, dans la configuration mobile, le locuteur se déplace lentement d'une position vers l'autre.

Dans notre base de données, on va étudier juste le cas de la configuration fixe comme illustrée dans la figure 3.3 :

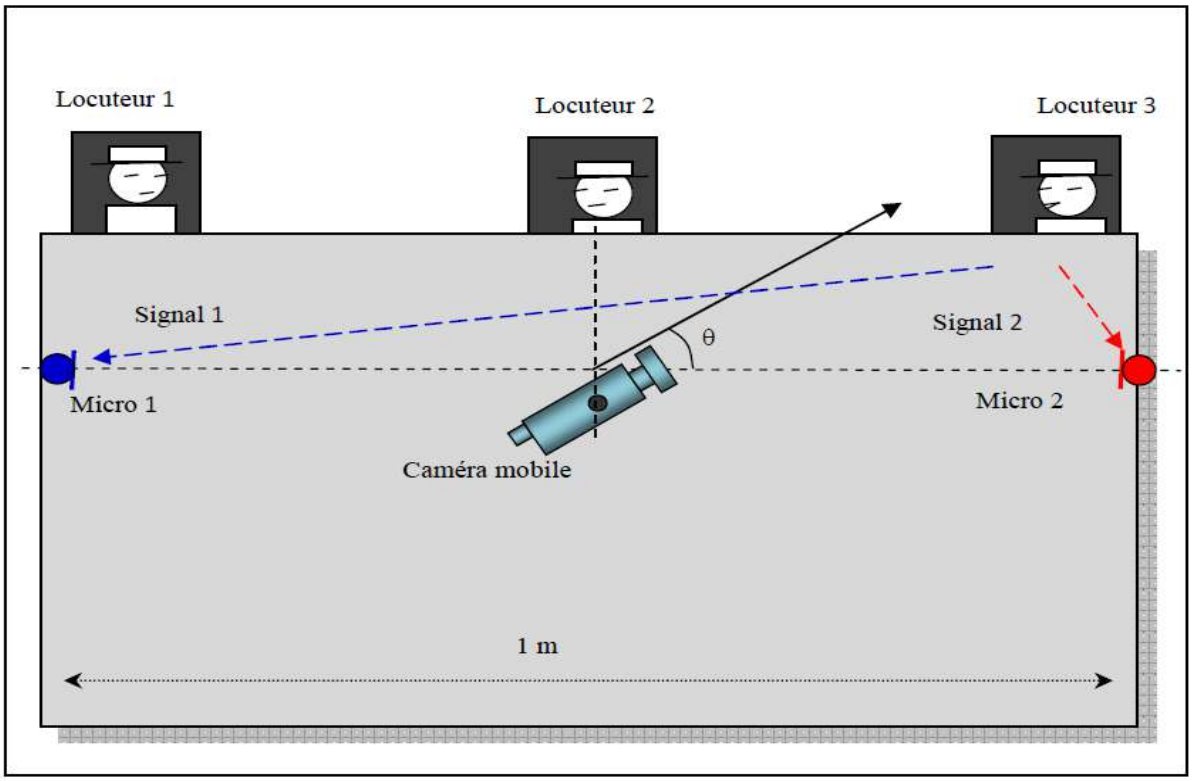


Figure3.3: Schéma synoptique de la configuration fixe.

III.2.4. Discours enregistrés

On a utilisés une variété de discours pendant les enregistrements audio de la base de donnéestels que : des sons phonémiques, des paragraphes, des discours et des conversations de manière naturelle.

III.4- Dispositifs matériels et logiciels utilisés

III.4.1- Schéma synoptique du système

Dans notre système, On distingue trois blocs essentiels: bloc d'acquisition, bloc de traitement et de commande et bloc d'exécution.

Le schéma synoptique du système automatique réalisé est illustré dans la figure 3.4:

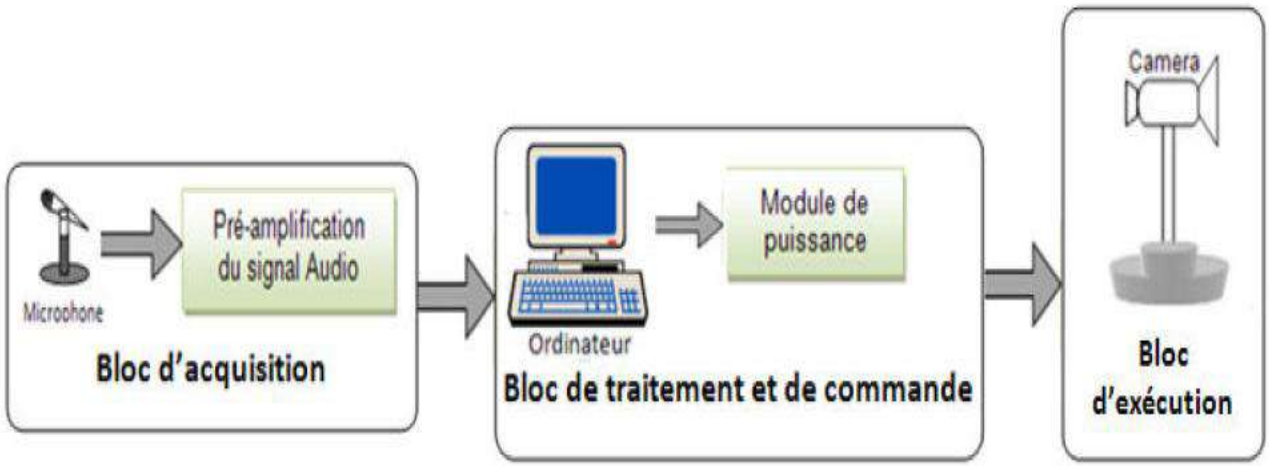


Figure 3.4: Schéma synoptique du système automatique réalisé.

III.4.2- Description des différents blocs du système

a) Bloc d'acquisition

Le bloc d'acquisition est constitué de deux microphones et d'un pré-ampli audio.
Le rôle de ce bloc estd'acquérir les signaux audio et de les amplifier avant qu'ils soient introduits au micro-ordinateur.

a.1) Microphones

Les signaux audio constituant la base de données BD11 sont perçus en utilisant deux microphones. Le premier est positionné à droite et le deuxième à gauche.



a) Microphone gauche b) Microphone droit

Figure3.5 : Photo des microphones utilisés dans les enregistrements.

Ces deux microphones sont les éléments du capteur les plus importants, plus le micro est performant, plus les mesures réalisées sur ce dernier sont précises. Ils possèdent les caractéristiques suivantes :

- Model dynamique unidirectionnel (cardioïdes).
- Portée de 2m maximale.
- Une réponsefréquentielle de 50Hz a 15000Hz.
- Une impédance de 150 Ohm.

a.2) Préampli audio

Les signaux audio enregistrés sont introduit par la suite dans unpré-amplificateur audio afin de les amplifier.
Le préampli utilisé est un préampli double voies, parfaitement adapté pour les microphones stéréo.
Il est basé sur l'emploi d'un circuit intégré de type NE5532. Ce circuit intégré comporte deux amplificateurs opérationnels au sein d'un mêmeboitier de huit broches, et qui présente de bonnes performances, notamment en termes de bruit. Il possède les caractéristiques suivantes:

- Impédance d'entrée : Environ 47KΩ pour chaque voie.
- Tension d'alimentation : +12 V.
- Gain du montage : 33dB.

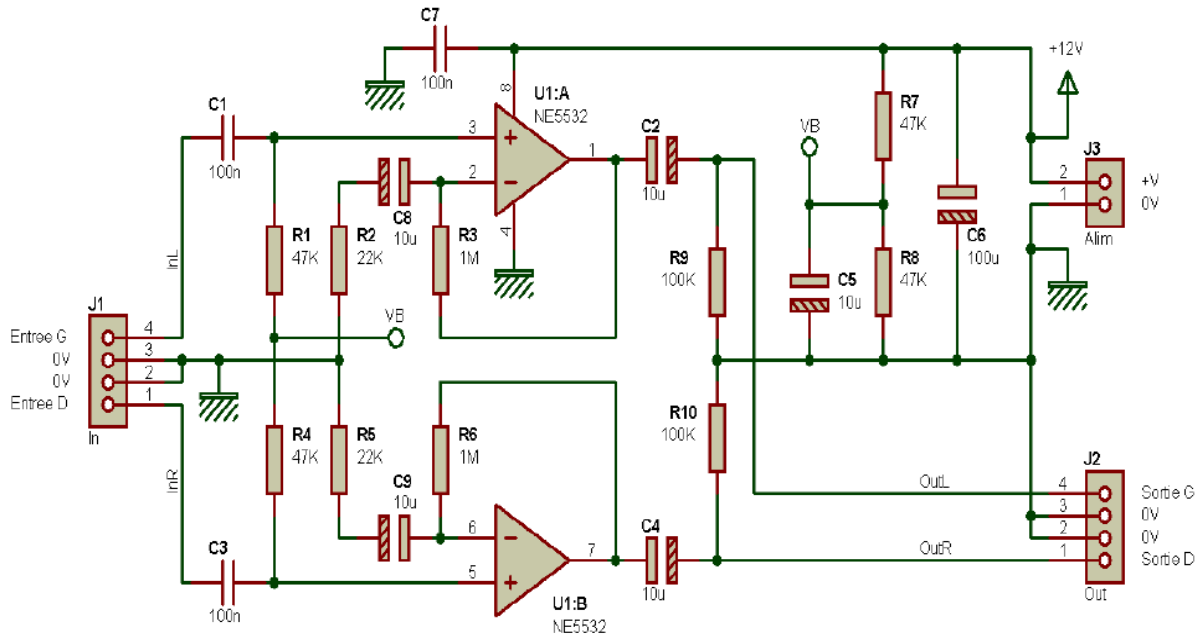


Figure 3.6: Schéma électrique du Pré-ampli audio utilisé.

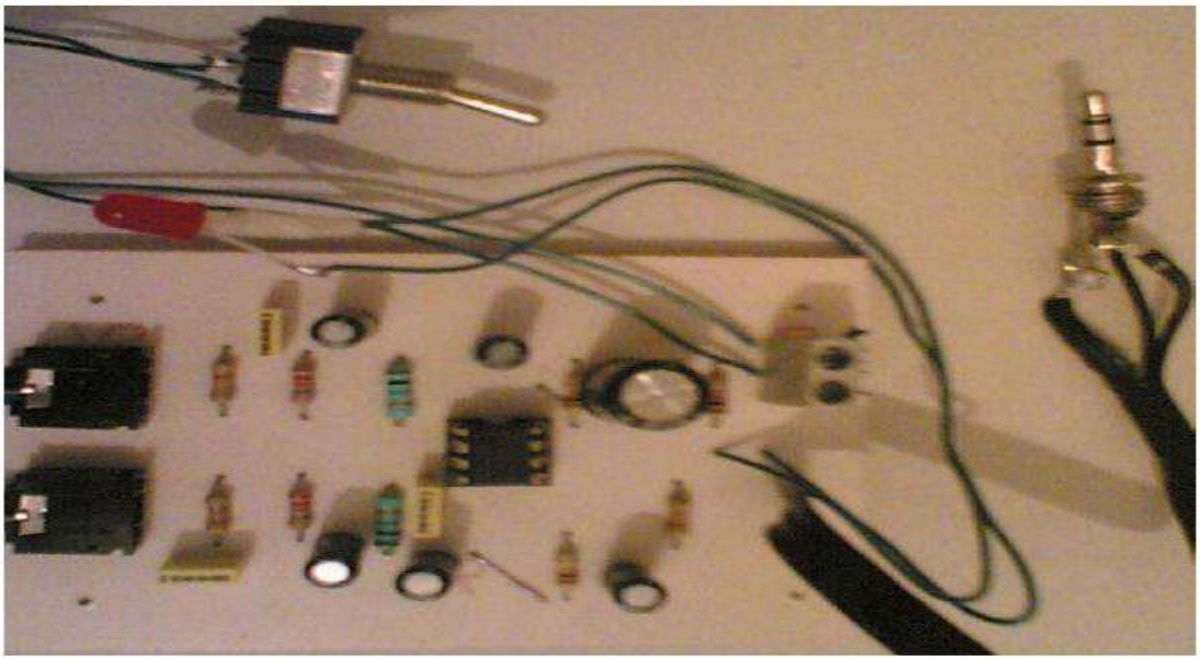


Figure3.7: Pré-ampli audio utilisé.

b) Bloc de traitement et de commande

Le rôle principal de ce bloc est de traiter les signaux audio enregistrés et de générer les signaux de commande adéquats. Il est constitué d'une unité de traitement et d'un circuit de puissance.

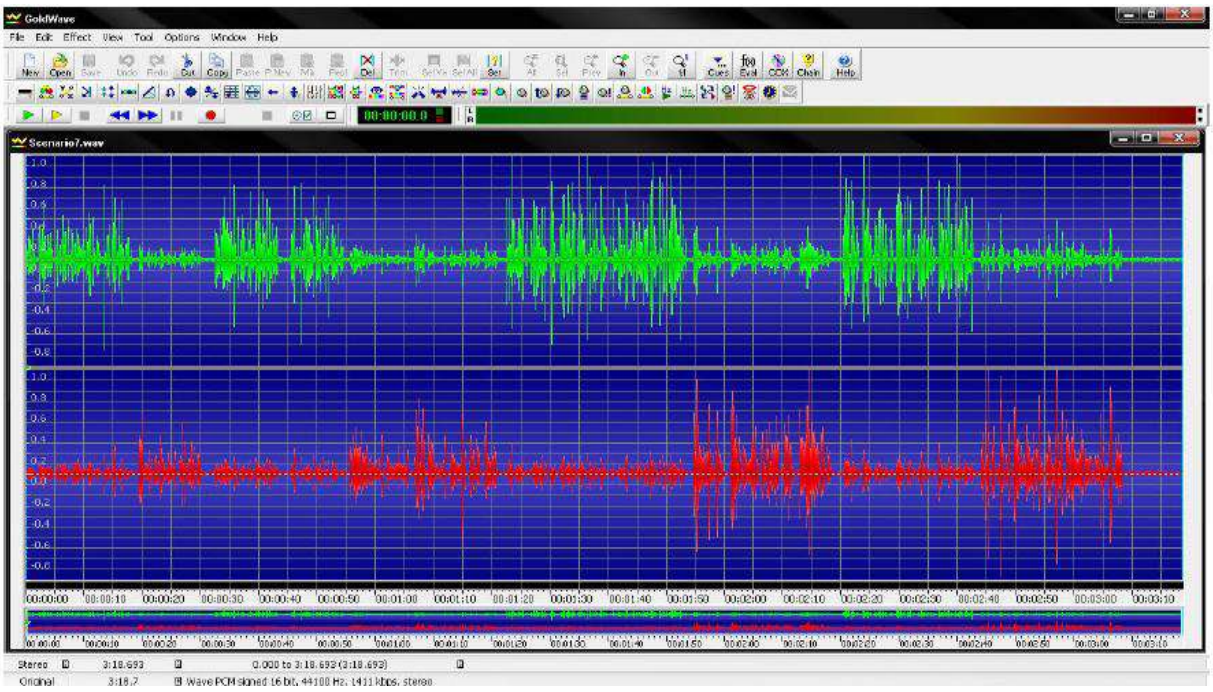
b.1) Unité de traitement

Elletraite les signaux perçus par les deux microphones (droit et gauche) et assure la commande du système via le port parallèle.
Cette unité de traitement ce n'est qu'un micro-ordinateur équipé de deux logiciels évolués :

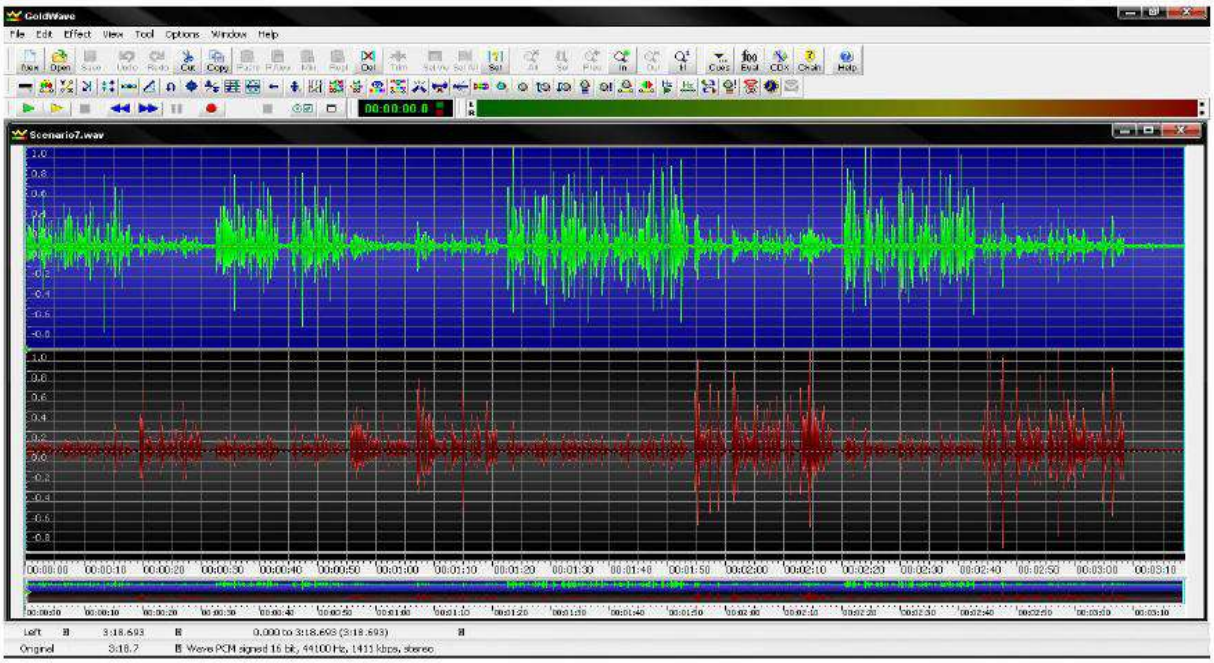
- Le Gold Wave.
- Le Matlab.

➢ **Pré-traitement audio avec le logiciel Gold Wave**

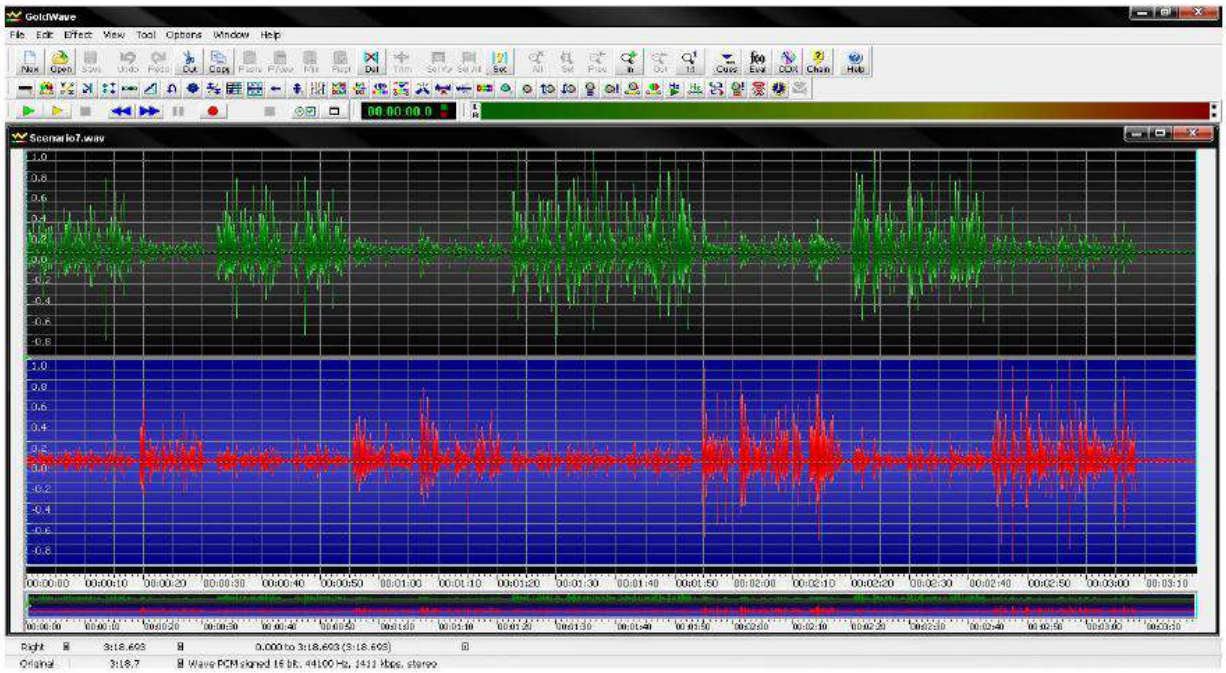
La dernière étape de préparation des signaux audio, qui sont des signaux stéréo, avant leur utilisation par notre système est le pré-traitement. Ce pré-traitement se fait par l'extraction des signaux de la voie gauche et de la voie droite du signal stéréo et les enregistrés sous des noms comportant l'indication de la voie concernée. Pour ce faire on a utilisé un logiciel appelé GoldWave.
Un exemple de ce pré-traitement est illustré par les photos présentées dans la figure suivante :



a) Signal stéréo (voie gauche et voie droite).



b) Sélection de la voie gauche.



c) Sélection de la voie droite.

Figure 3.8: Photo montrant le pré-traitement en utilisant le Gold Wave.

- **Traitement audio avec le logiciel Matlab**

Matlab est l'outil de référence pour la simulation numérique, notamment en ce qui concerne l'automatique et la commande. Il permet, de manière plus générale, de résoudre une grande diversité deproblèmes de simulation, dans des domaines aussi variés que le traitement du signal, les statistiques ou la vision.
C'est sur Matlab qu'on a essayé de faire nos programmes et de tester les performances des différentes méthodes de localisation proposées.

b.2) Circuit de puissance

L'interfaçage entre notre systèmeautomatique et l'ordinateur via le port parallèle est assuré par un circuit de puissance.

c) Bloc d'exécution

Lerôle de ce bloc est d'assurer l'orientation d'une caméra vers la direction appropriée.
A l'aide du logiciel Matlab, l'acquisition des images peut être faite directement à partir de la caméra et la transmettre à la carte graphique du micro-ordinateur.

III.5- Conclusion

Dans ce chapitre, nous avons présenté les détails de conception de notre base de données BD11 ainsi que lesdispositifs matériels et logiciels utilisés.
Avec cette base de données, on a obtenu des résultats satisfaisants en utilisant les méthodes de localisation proposées dans le chapitre précédant.

IV.1- Introduction

Dans ce chapitre, on va exposer les résultats des expériences effectuées en utilisant différentes méthodes de localisation appliquées sur plusieurs scénarios de parole. le logiciel utilisé est le Matlab.

Ces résultats seront interprétés et comparés d'une manière adéquate afin d'extraire les performances de chaque méthode quant à la qualité des résultats de localisation.

IV.2- Etapes des expériences réalisées

IV.2.1- Configuration de la base de données BD11

Notre base de données BD11 constituée de 6 scénarios, contenant des locuteurs masculins et féminins qui parlent de manière les naturelle, ces derniers sont assis dans des positions fixes. les expériences de localisation consistent à détecter la position du locuteur actif (qui parle) et ceci par le biais de sa parole.

IV.2.2- Acquisition des signaux audio

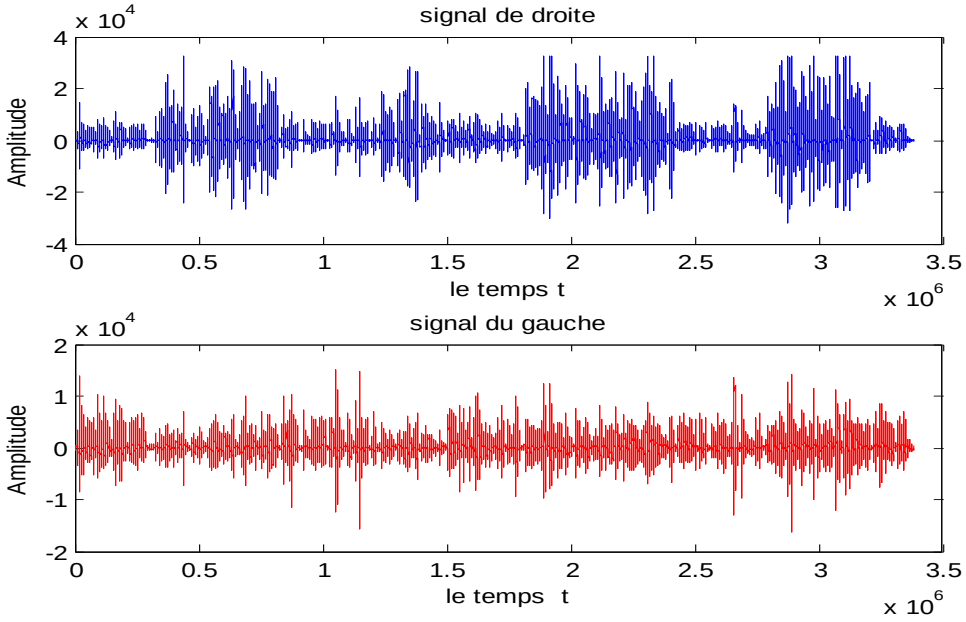
Les signaux audio sont enregistrés par deux microphones (gauche et droite), et amplifiés par un pré-ampli. les deux microphones sont positionné en opposition et séparés par une distance fixe (1mètre)

IV.2.3- Pré-traitement audio avec le logiciel Gold Wave

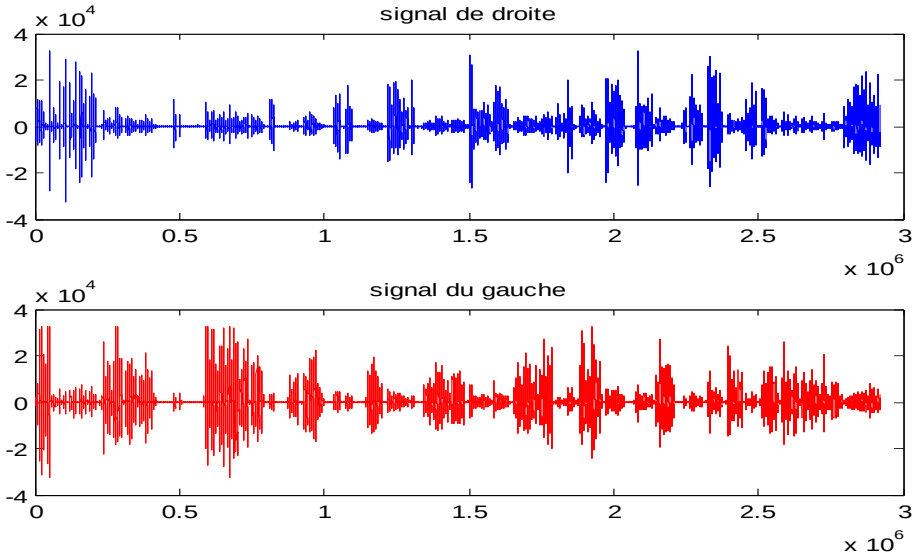
L'écoute est le premier moyen qui permet d'apprécier la qualité d'un signal, et de détecter les éventuels problèmes créées par le système de prise de son (Akchiche, 2007).

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

La figure 4.1est un exemple de lecture des signaux audio par le logiciel Gold Wave.



- a -



- b -

Figure 4.1: Visualisation des signaux audio.

a) Signaux issus d'une discussion entre deux locuteurs dont l'un est au milieu et l'autre est à gauche (scénario 1 de la BD11).

b) Signaux issus d'une discussion entre deux locuteurs dont l'un est à droite et l'autre est à gauche (scénario 5 de la BD11).

IV.2.4- Traitement audio avec le logiciel Matlab

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

- Méthode du différentiel d'énergie (ou en anglais Energy Differential Method **EDM**).
- Méthode de Generalized Cross-Correlation PHAse Transform (**GCC –PHAT**).

A fin d'obtenir une précision fine 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 représentés comme suit : la gauche (-1), le milieu (0) et la droite (1).

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

a) Méthode du différentiel d'énergie EDM

Avec cette méthode, la position est détectée selon la valeur de DE (différence d'énergie) qui est le différentiel de l'énergie logarithmique donné par la relation (2.4) (du chapitre 2).

Cette différence d'énergie DE sera comparée à un seuil de détection bien choisi expérimentalement.

b.1) Algorithme

L'algorithme suivi pour réaliser cette expérience est défini comme suit :

```
% la lecture des signaux issus respectivement du micro 'd' et du
micro 'g'

Lire (Sd)
Lire (Sg)

% fenêtrage des signaux

Décomposition des deux signaux de parole en fenêtres de taille 0.5 s
   Nombre de fenêtre = taille du signal/taille d'une fenêtre

%calcul de l'énergie des deux signaux et estimation de la position

 Pour   i=1:nombre de fenêtre

Calcul de l'énergie des deux signaux le long de la fenêtre

 Calculer (E_Sd (dB))
 Calculer (E_Sg (dB))

Calcul de différentiel d'énergie

 DE = E_Sd (dB) - E_Sg (dB)

Estimation de la position
     Si DE>seuil_max
                Position='droite'
     Sinon Si DE<seuil_min
                Position='gauche'
     Sinon
                Position='milieu'
     Fin Si;

Fin pour
```

b.2) Résultats des expériences réalisées

En appliquant l'algorithme précédent aux différents scénarios de la base de données BD11, on a obtenu les résultats illustrés dans les figures (4.2) à (4.7).

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

Scénario 1: Halim-Hakim_ (droite-milieu)

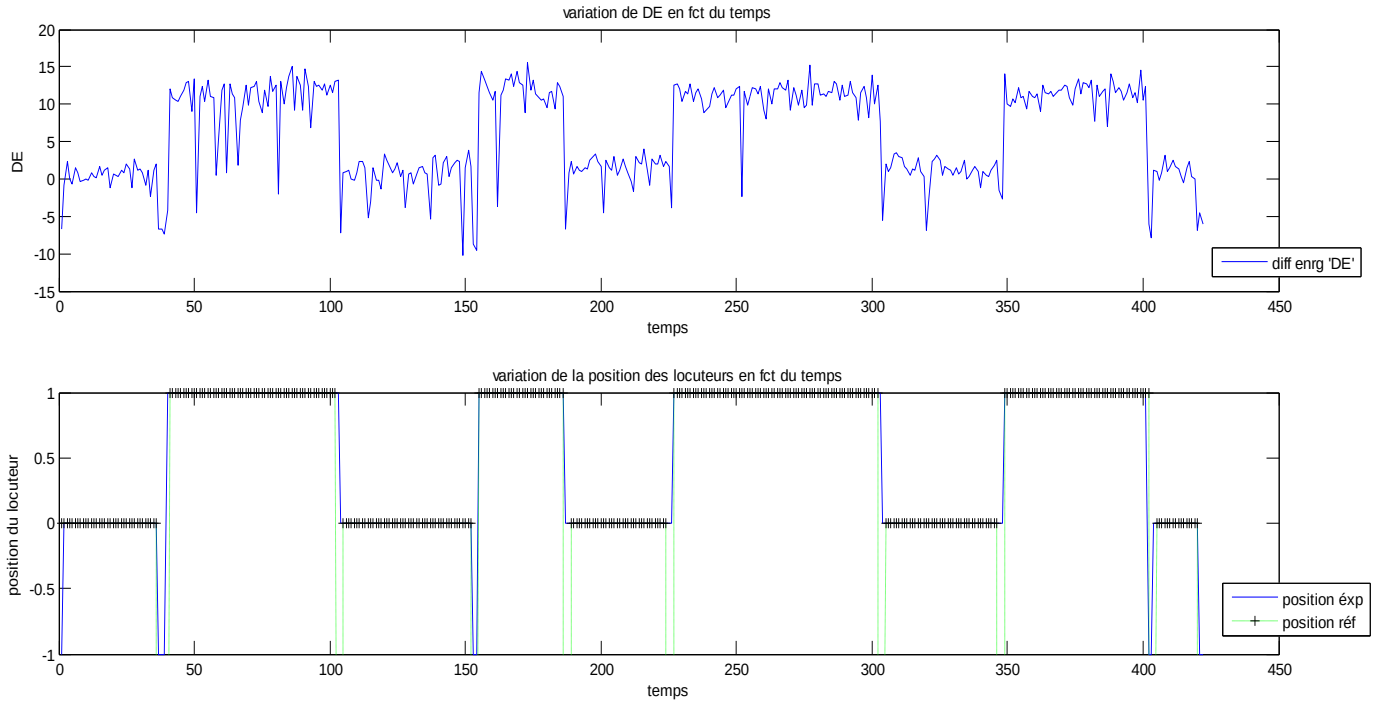


Figure 4.2 : Résultats de localisation obtenus par la méthode EDM pour le scénario 1.

Scénario 2: Halim-Siham_(milieu-droite)

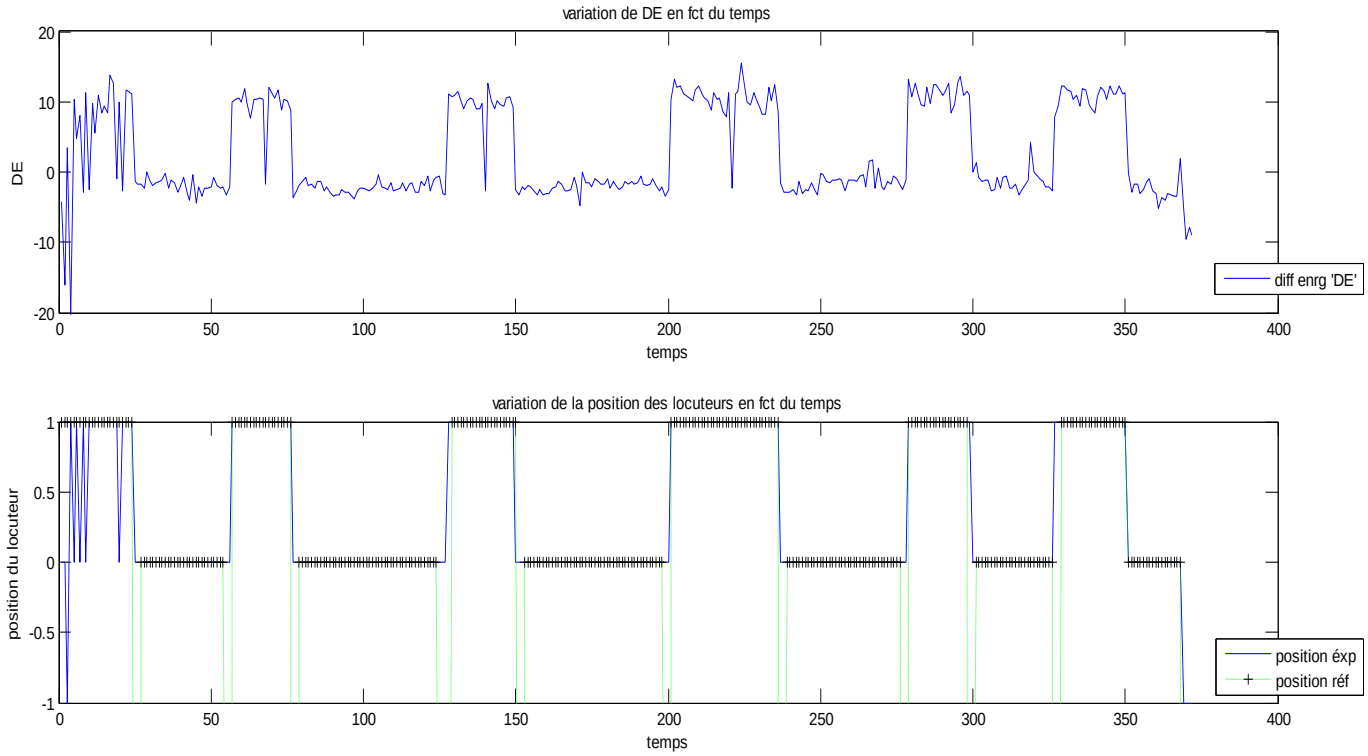


Figure 4.3 : Résultats de localisation obtenus par la méthode EDM pour le scénario 2.

Scénario 3: Nadia-Halim (gauche-milieu)

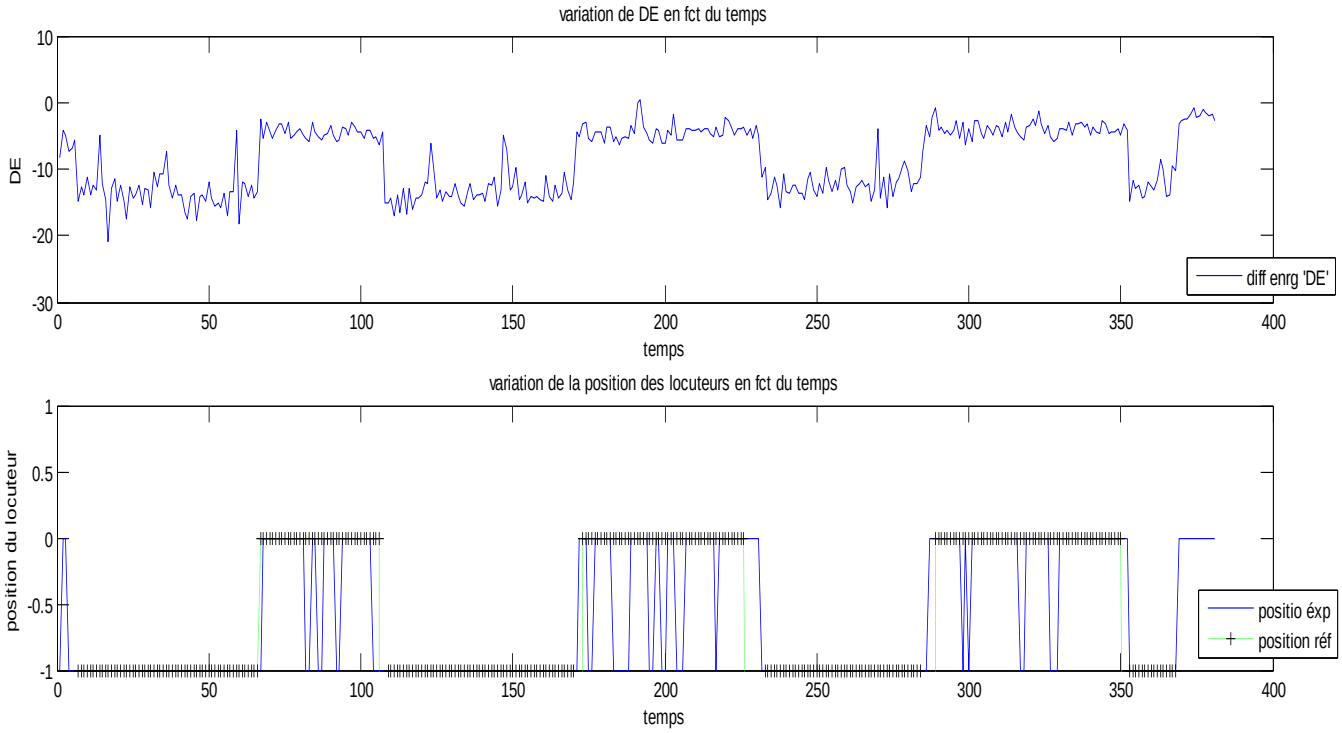


Figure 4.4 : Résultats de localisation obtenus par la méthode EDM pour le scénario 3.

Scénario 4 : Hakim-Halim_(gauche-droite)

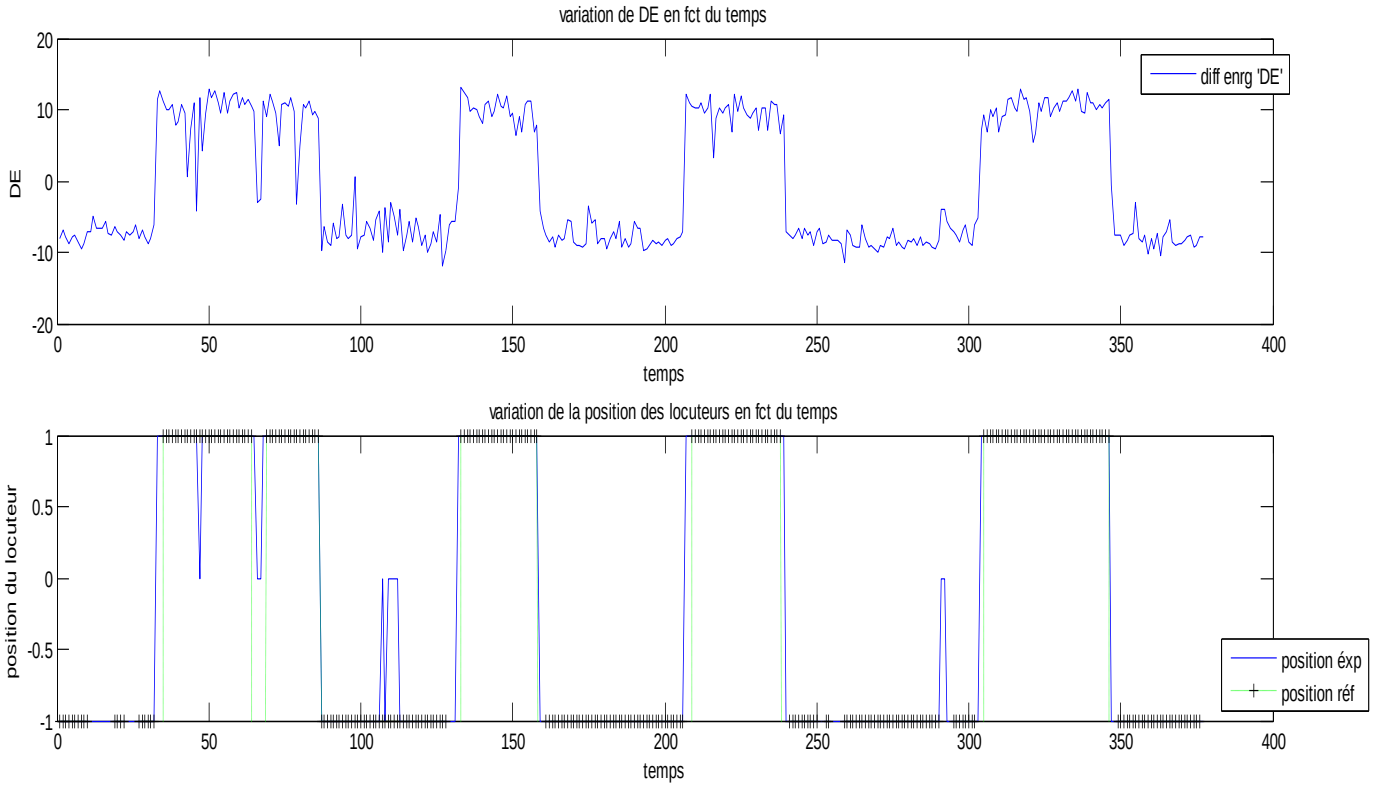


Figure 4.5 : Résultats de localisation obtenus par la méthode EDM pour le scénario 4.

Scénario 5:Halim-Siham_(gauche-droite)

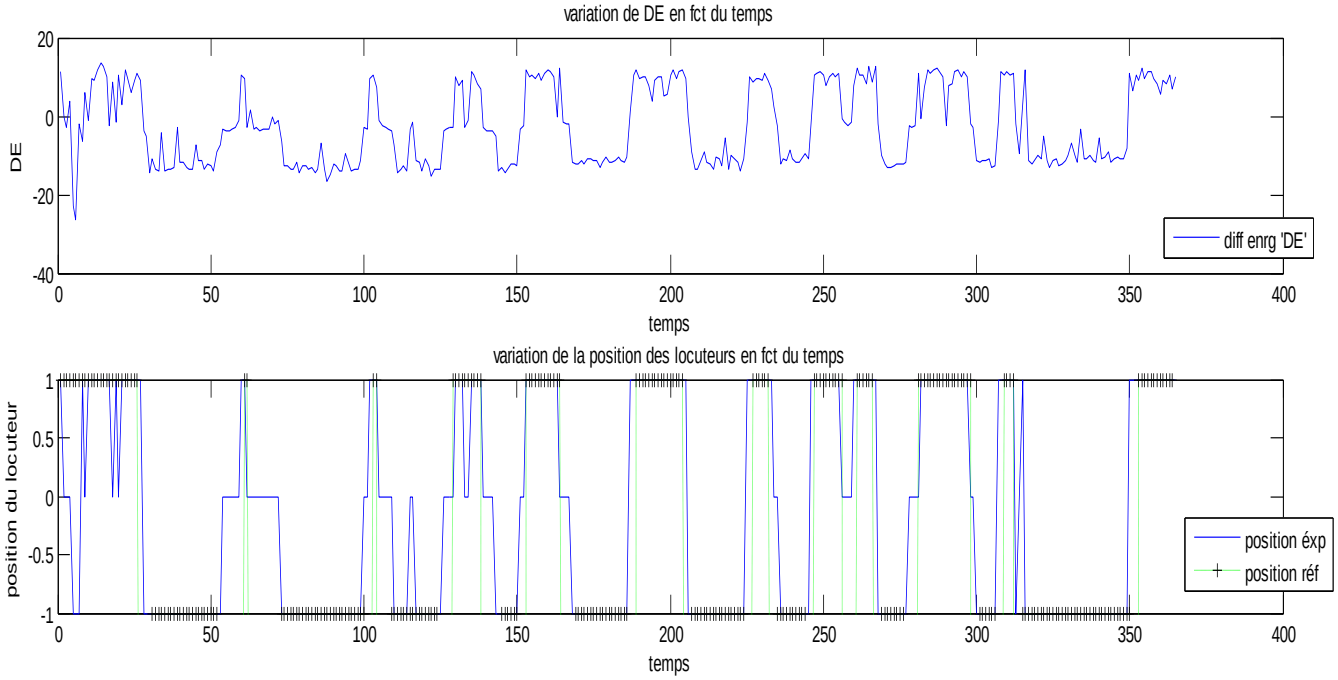


Figure 4.6 : Résultats de localisation obtenus par la méthode EDM pour le scénario 5.

Scénario 6:Nadia-Halim_ (gauche-droite)

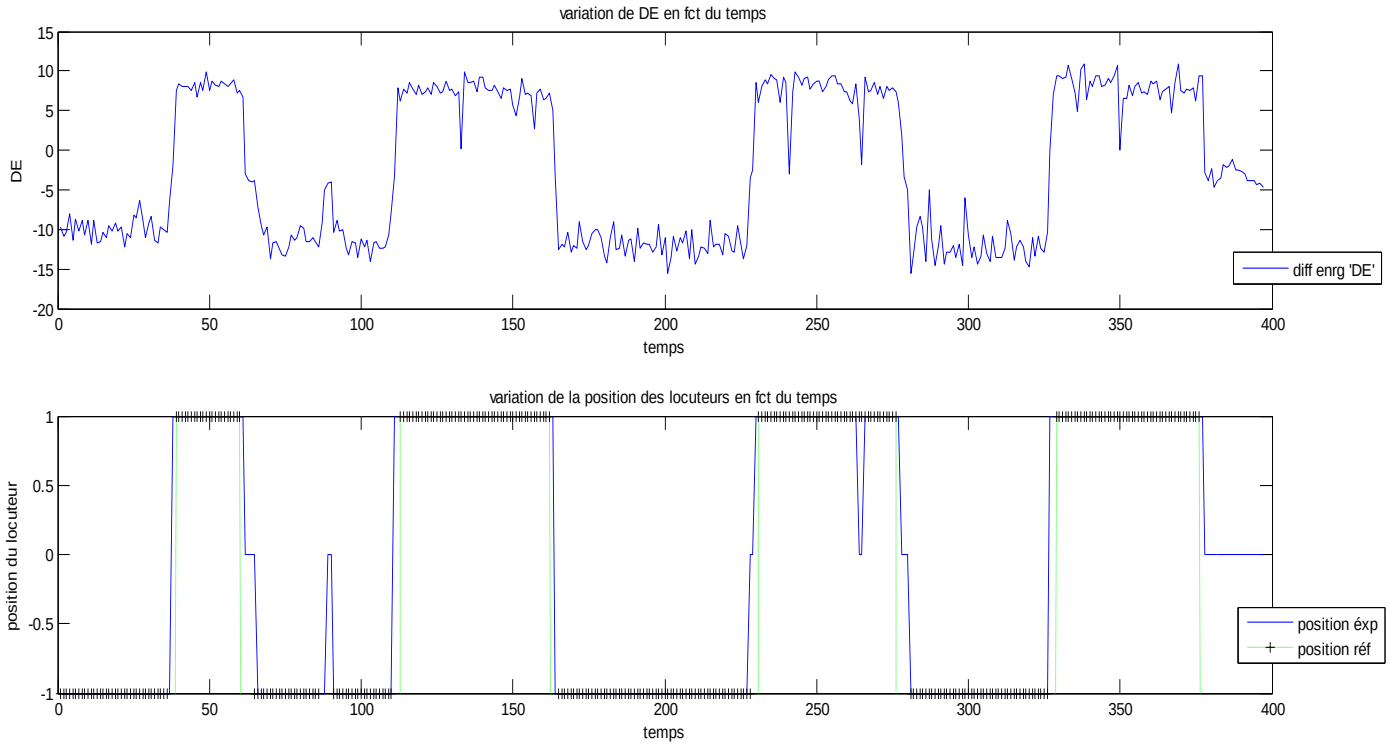


Figure 4.7 : Résultats de localisation obtenus par la méthode EDM pour le scénario 6.

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

Pour la méthode de la **GCC-PHAT** : on a utilisé la formule (2.27) du chapitre 2, qui nous permet de calculer le retard ($\boldsymbol{\tau}$) 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) donnée par la relation (2.8) du chapitre 2.
Cette distance sera comparée un seuil de détection bien choisi expérimentalement.

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

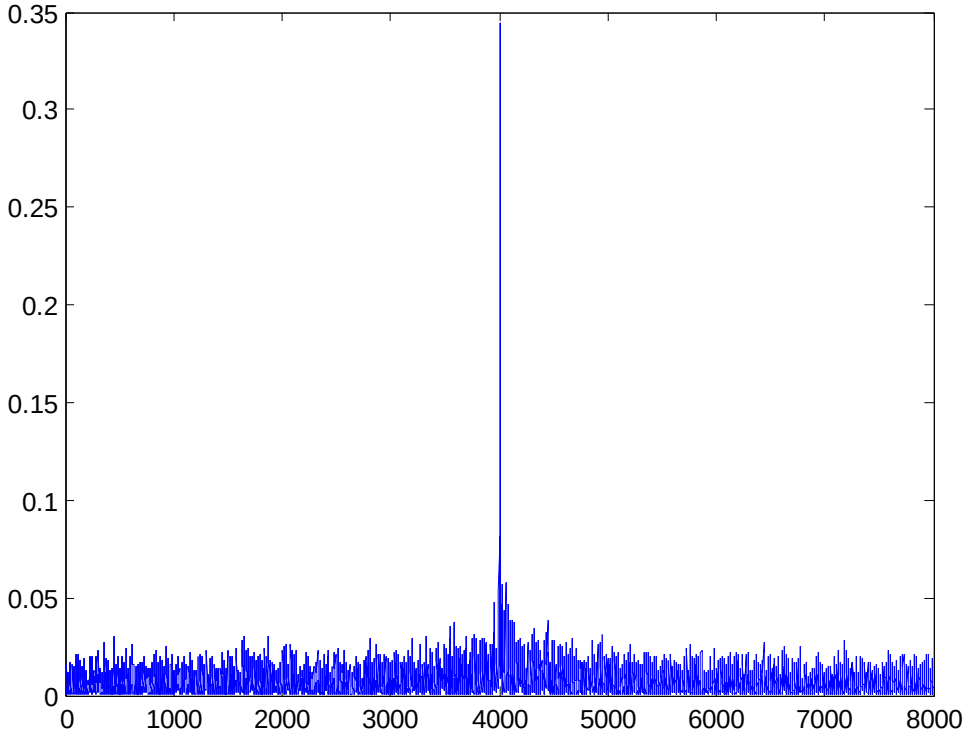


Figure 4.8 : exemple de localisation d'un locuteur au milieu (scénario 1)

- Le pic donne un retard $\tau = 2{,}5\ e^{-4}$ s (4.1)
  La distance **d** est : d=0,5426 m (0,4<d<0,6) (4.2)

Donc : le locuteur qui parle est au milieu.

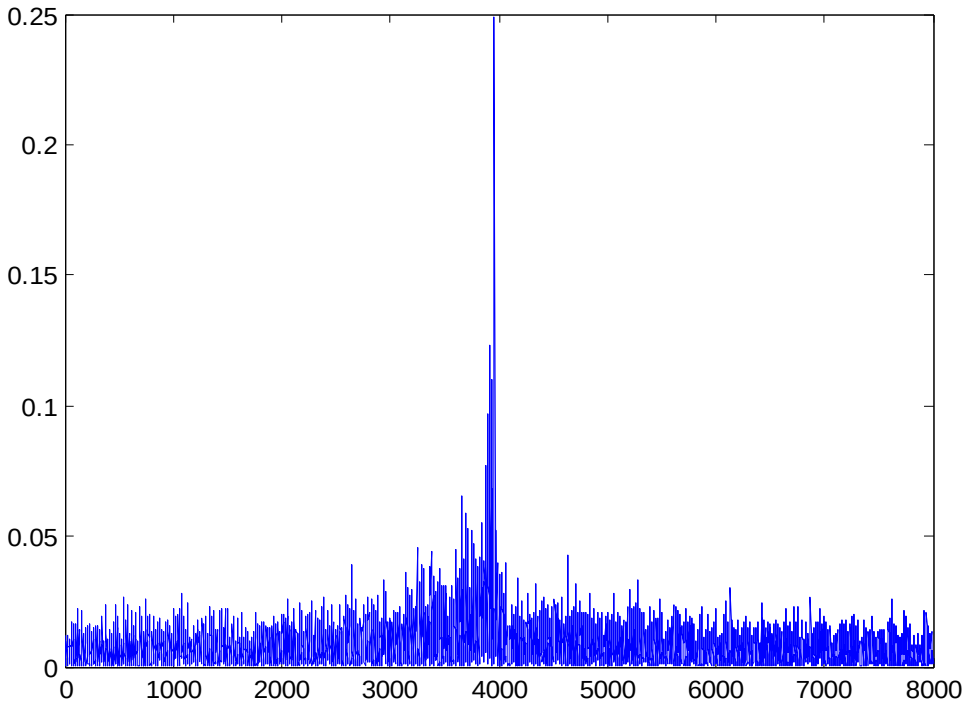


Figure 4.9 : exemple de localisation d'un locuteur à droite (scénario 1)

- Le pic donne un retard $\tau = -0.0024s$ (4.2)
  La distance **d** est : d=0,092 m (d<0,4) (4.3)

Donc : le locuteur qui parle est à droite.

c.1) Algorithme

Voici l'algorithme suivi pour réaliser cette expérience :

```
% la lecture des signaux issus respectivement du micro 'd' et du
micro 'g'

Lire (Sd)
Lire (Sg)

% fenêtrage des signaux

Décomposition des deux signaux de parole en fenêtres de taille 0.5 s
   Nombre de fenêtre=taille du signal/taille d'une fenêtre

%calcul du retard (τ) et estimation de la position

 Pour   i=1:nombre de fenêtre

Calcul de la GCC-PHAT sur chaque fenêtre
```

$$GCC - PHAT = ifft\left(\frac{A.B^*}{|A|.|B|}\right)$$

```
Détection de la position du maximum, donc (τ)
```

$$\tau = argmax\,(GCC - PHAT)$$

```
          Calcul de la distance entre les locuteurs et les microphones
```

$$d=(L+\tau*Vs)/2;$$

```
Estimation de la position
     Si d>seuil_max
                Position='droite'
     Sinon Si d<seuil_min
                Position='gauche'
     Sinon
                Position='milieu'
     Fin Si;
Fin pour
```

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

Scénario 1:Halim-Hakim_ (droite-milieu)

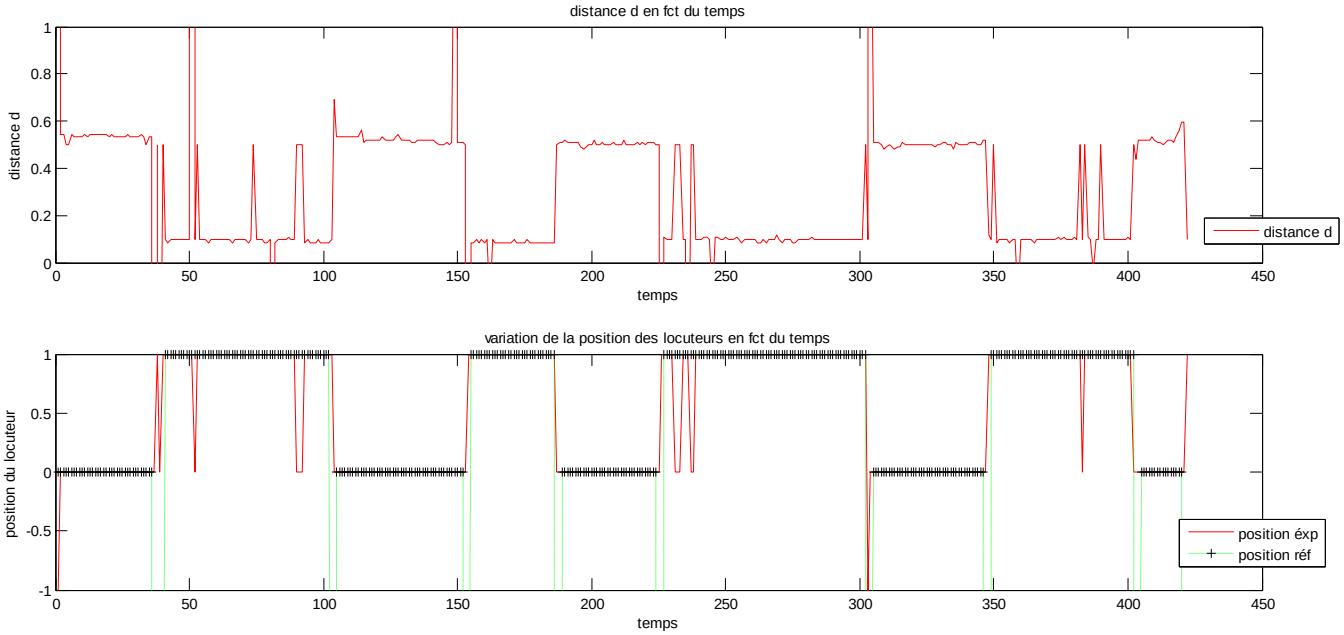


Figure 4.10: Résultats de localisation obtenus par la méthode GCC-PHAT pour le scénario 1.

Scénario 2:Halim-Siham_(milieu-droite)

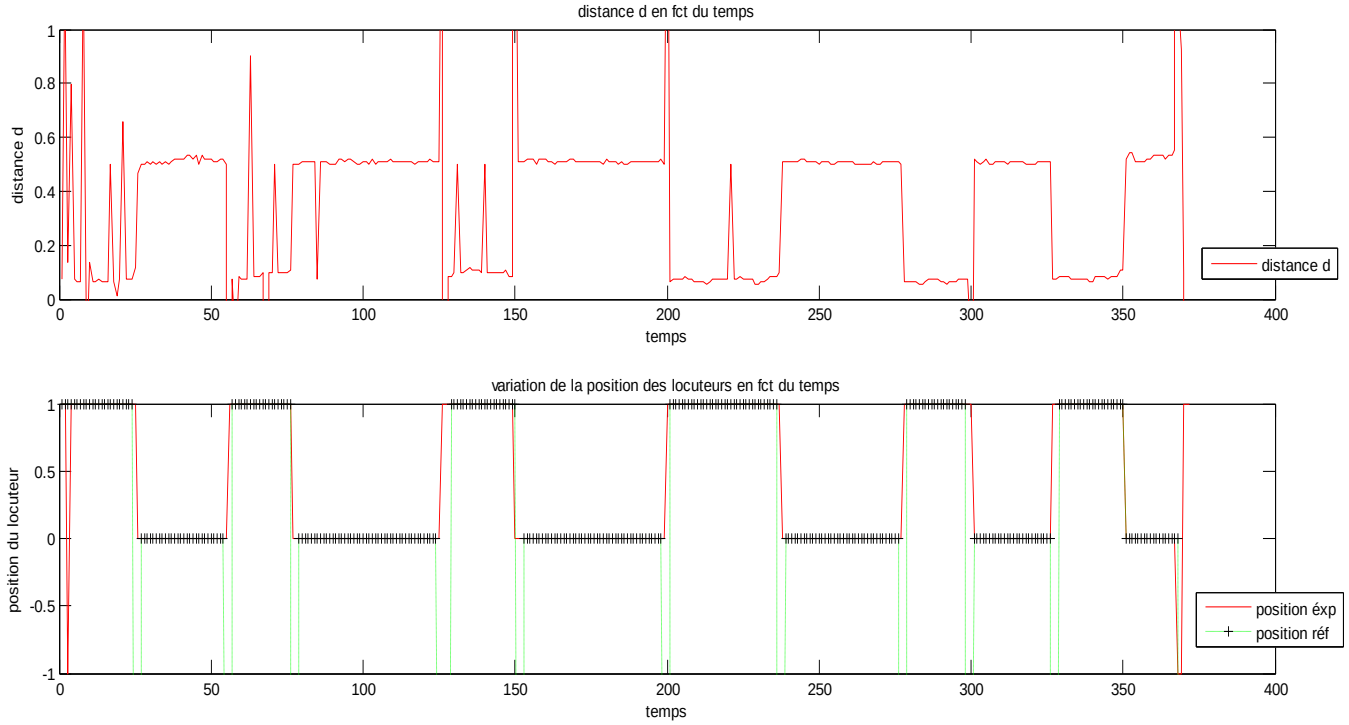


Figure 4.11 : Résultats de localisation obtenus par la méthode GCC-PHAT pour le scénario 2.

Scénario 3:Nadia-Halim_ (gauche-milieu)

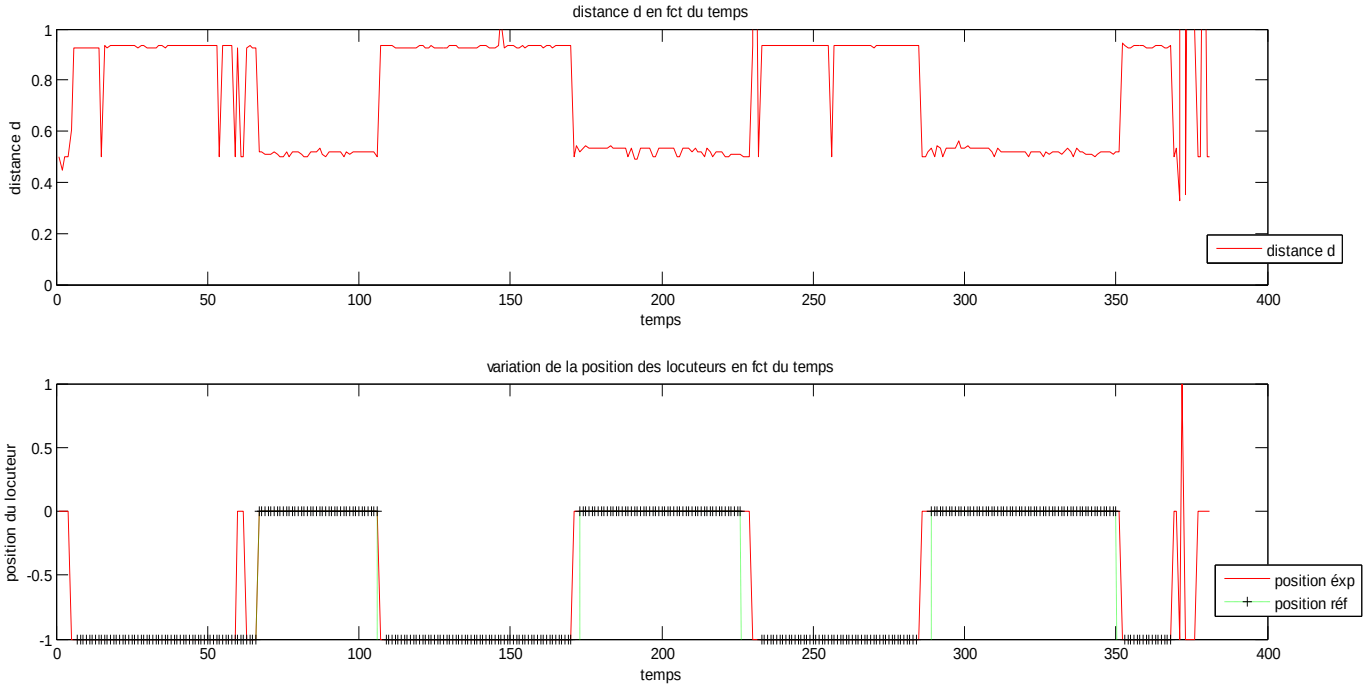


Figure 4.12 : Résultats de localisation obtenus par la méthode GCC-PHAT pour le scénario 3.

Scénario 4:Hakim-Halim_(gauche-droite)

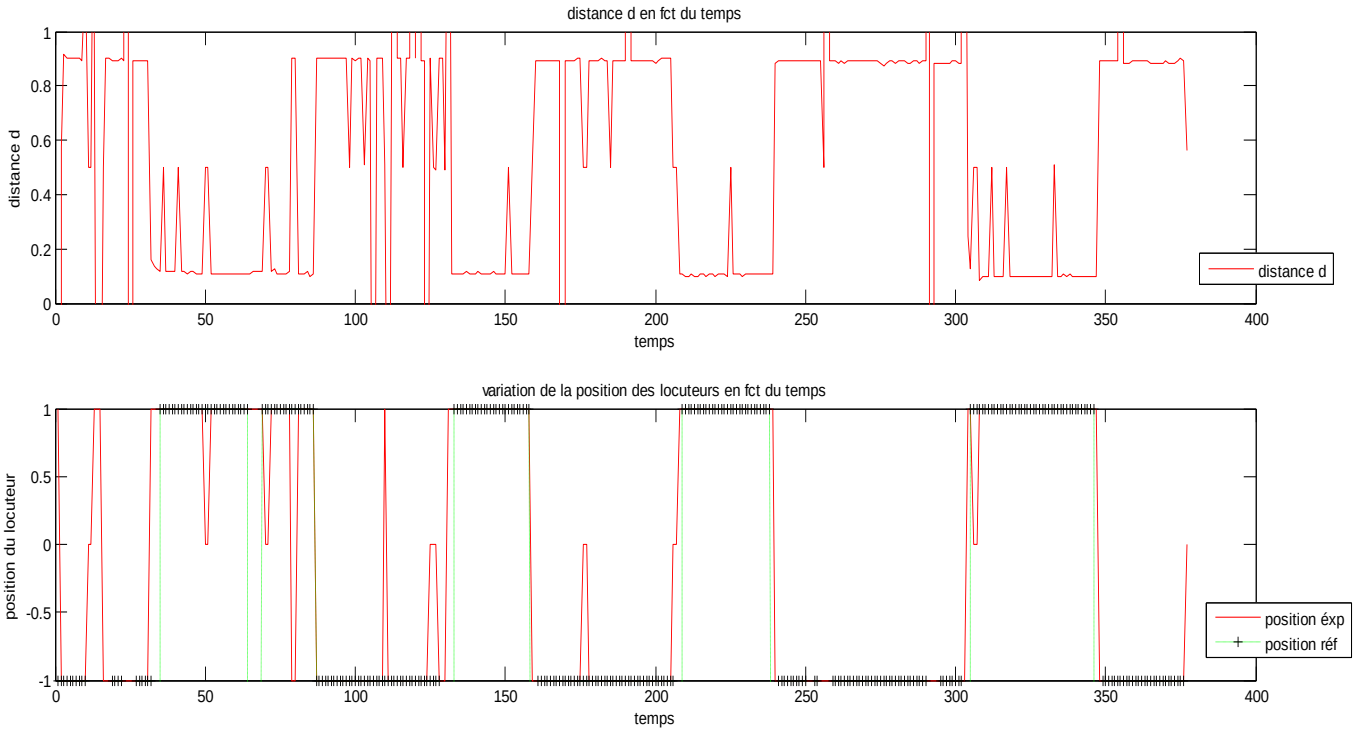


Figure 4.13 : Résultats de localisation obtenus par la méthode GCC-PHAT pour le scénario 4.

Scénario 5:Halim-Siham_(gauche-droite)

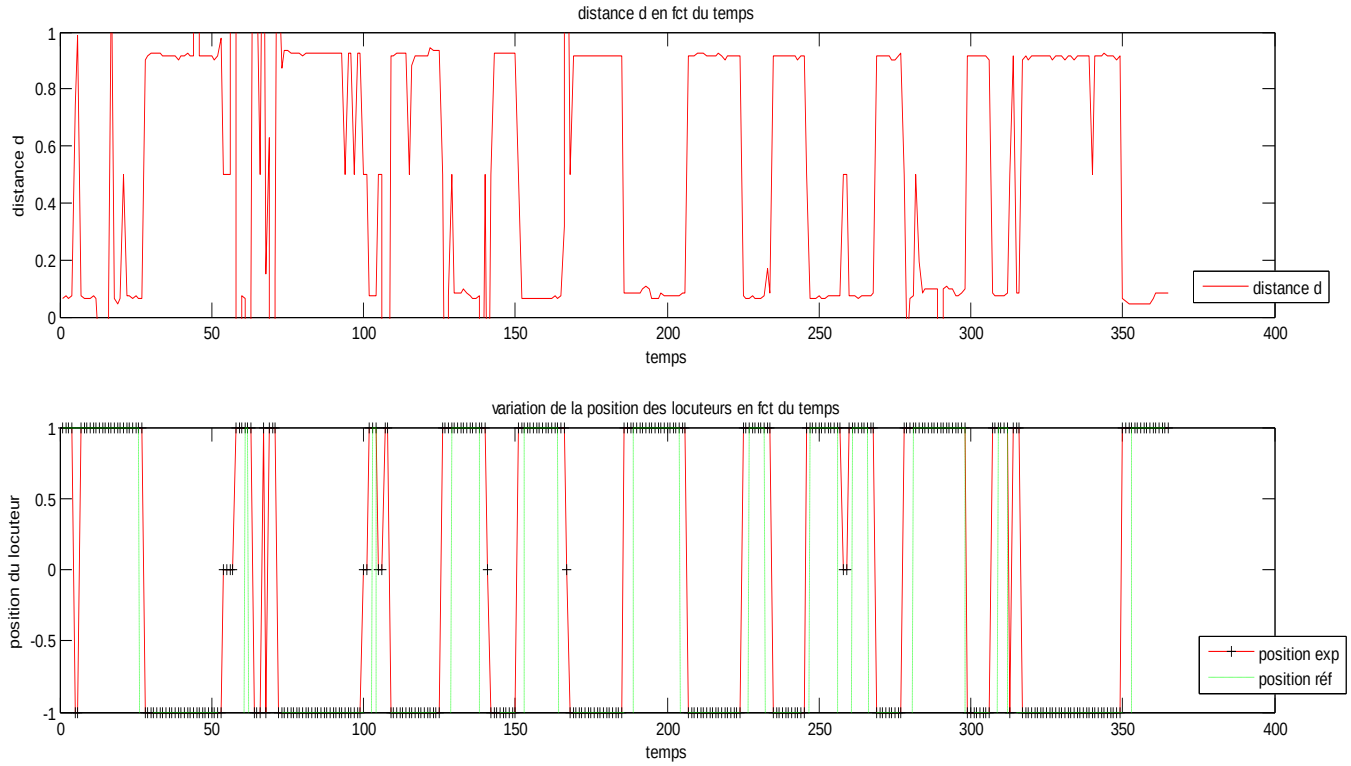


Figure 4.14 : Résultats de localisation obtenus par la méthode GCC-PHAT pour le scénario 5.

Scénario 6:Nadia-Halim_(gauche-droite)

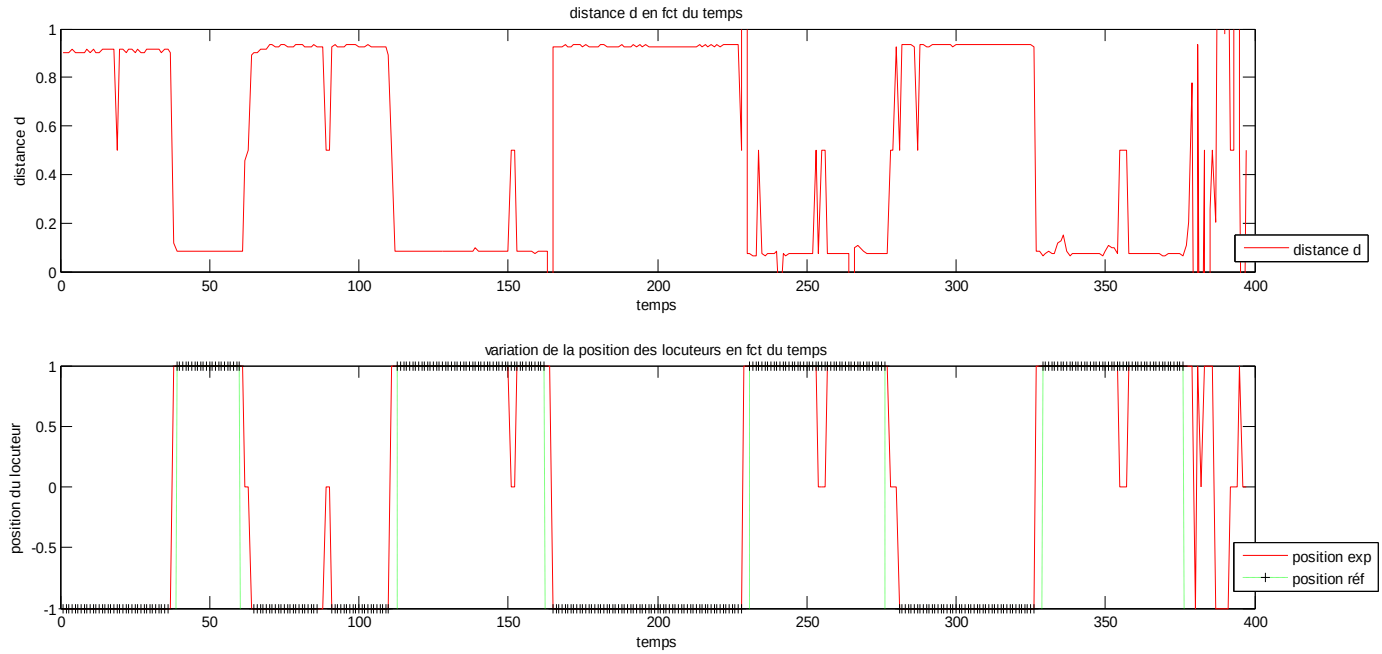


Figure 4.15 : Résultats de localisation obtenus par la méthode GCC-PHAT pour le scénario 6.

IV.3- Statistiques des résultats obtenus

La simulation avec Matlab a donné les résultats illustrés dans le tableau ci-dessous:

Tableau 4.1: Taux de précision de localisation de la position en % par scénario.

techniques / scénarios	EDM	GCC-PHAT
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**

Avec les résultats, on a tracé l'histogramme suivant:

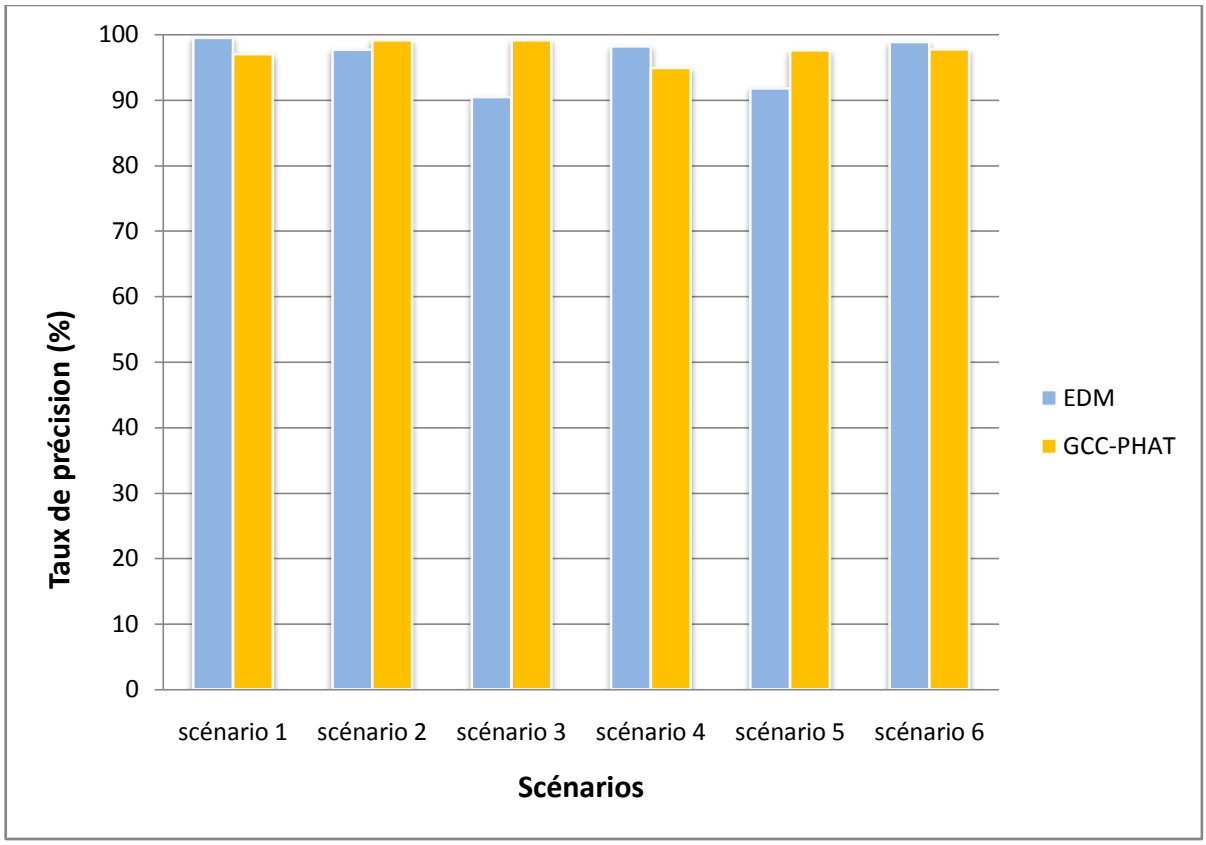


Figure 4.16: Taux de précision de localisation de la position en % par scénario.

Commentaire

- La méthode **EDM** présente une variation de taux de précision de la localisation entre 90,46 % et 99,5 %. Ces valeurs changent d'un scénario à un autre.

- La méthode **GCC-PHAT** présente une variation de taux de précision de la localisation entre 94,91 % et 99,13 %. Ces valeurs changent d'un scénario à un autre.

Pour pouvoir comparer les performances des deux méthodes, il est préférable de calculer la moyenne de taux de précision de chacune d'elle.

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

Tableau 4.2: La moyenne des taux de précision par méthode.

EDM	GCC-PHAT
96 ,03 %	**97,59 %**

Les résultats obtenus sous forme d'histogramme est dans la figure suivante :

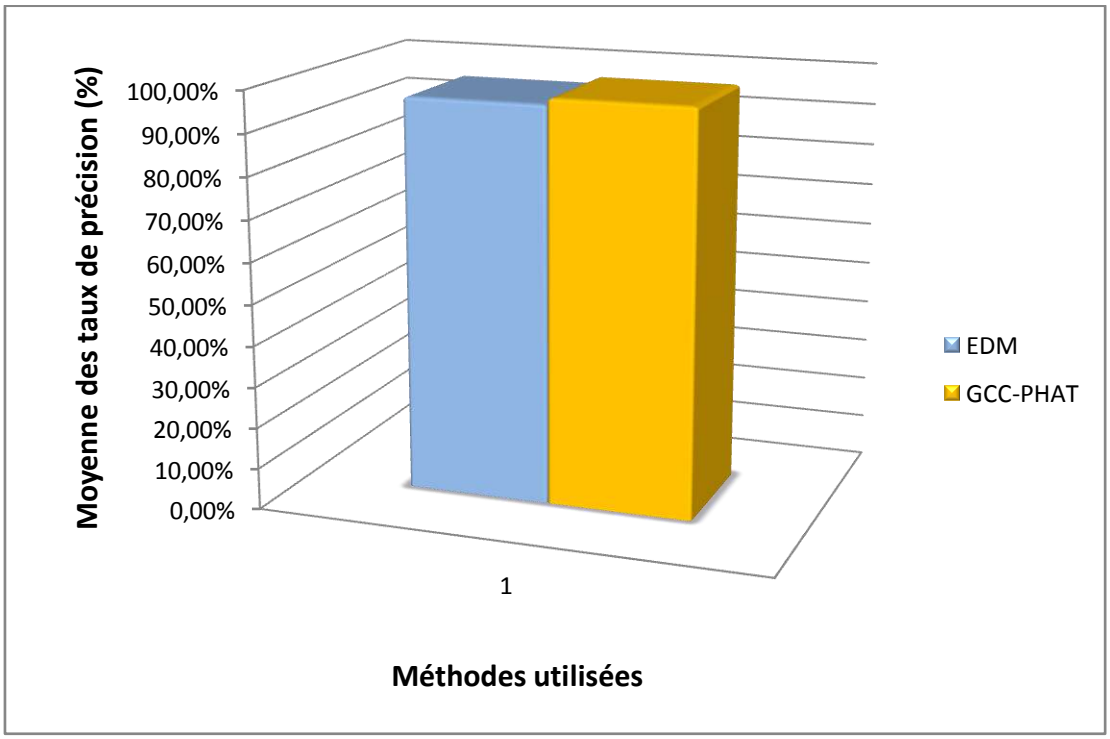


Figure 4.17: La moyenne des taux de précision pour chaque méthode.

Commentaire :

Les deux méthodes **EDM** et **GCC-PHAT** donnent des résultats proches et satisfaisants concernant le taux de précision de la localisation mais la **GCC-PHAT (97,59 %)** était un peu plus meilleure car elle est très proche de l'objectif voulu**.**

IV.4- Conclusion

Dans ce chapitre, on a pu montrer l'efficacité des trois techniques utilisées dans la détection spatiale des locuteurs avec des signaux stéréophoniques en utilisant une paire de microphones.

Parmi ces techniques, la CCG-PHAT était la plus performante avec un taux de précision qui atteint les **97,59 %.**

Comme nous l'avons mentionné dans l'introduction de ce manuscrit, le but de notre travail est d'effectuer une étude sur la localisation des sources sonores présentée par des locuteurs qui parlent à l'aide des signaux stéréophoniques.
Notre application était destinée à la supervision automatique des enregistrements audiovisuels dans les meeting-rooms dans le but d'améliorer les performances des systèmes de détection du son.

Au début nous avons présenté les différentes méthodes de détection telle que: le différentiel d'énergie **(EDM)** et la Generalized cross-correlation phase transform (**GCC –PHAT**).

La plupart de ces méthodes a fait l'objet de plusieurs études comparatives qui ont montré les avantages des unes et les inconvénients des autres.

Et pour pouvoir juger d'une manière adéquate nos résultats, plusieurs scénarios ont été prise en compte, parfois, on change les locuteurs (masculins ou féminins) et d'autre fois, on change leurs positions (gauche, milieu ou droite).

Discussion des résultats obtenus :

Les résultats expérimentaux obtenus ont montrés :

- Une bonne précision de localisation oùla moyenne des deux techniques dedétection atteint environ les **96,81 %**de bonne détection.

- Une meilleure précision de détection pour la méthode basée sur la GCC-PHATqui atteint environ les **97,59 %**surcelle qui utilise le différentiel d'énergieEDM.

Ces résultats sont obtenus pour des positions fixes et en utilisant un nombre limité de locuteurs (3 personnes seulement) ainsi on a traité uniquement deux techniques de localisation (EDM et GCC-PHAT) mais il sera intéressant d'augmenter le nombre de locuteur (gauche, droite, milieu, milieu gauche et milieu droite) et appliquer ces méthodes sur des locuteurs mobiles ainsi d'utiliser d'autres techniques de localisation.
De plus, les erreurs de localisation peuvent être réduites d'avantages si on renforce les conditions d'enregistrement audio par une réduction des échos (chambre sourde) ou par l'utilisation de microphones de meilleure directivité tel que les microphones matricielles (microphone arrays) qui nous offre la possibilité de localiser les locuteurs dans le plan (X,Y) et dans l'espace (X,Y,Z).

Ceci n'empêche pas de dire qu'à travers cette étude, nous estimons que le projet que nous avons mené nous a été fort utile de par les connaissances théoriques mais surtout pratiques que nous avons acquis.

- ***Les documents écrits :***

(Akchiche**, 2007)** Akchiche S.,"Identification du locuteur: Approche par quantification vectorielle", Mémoire d'ingéniorat en génie électrique, Ecole Militaire Polytechnique, 2007.

(Ait Saadi,**2005**) Ait Saadi N.,Takhedmit H., "Identification du locuteur en mode indépendant du texte", Mémoire de Doctorat en électronique, Ecole Nationale Polytechnique, El Harrach, Alger, Juin 2005.

(Benselama,**2007**) BenselamaZoubir-Abdeslem.,"Pathologie du Langage Parlé Arabe Cas des Sigmatismes Occlusifs et Constrictifs", Mémoire de Doctorat d'état en électronique, Ecole Nationale Polytechnique, El Harrach, Alger, Décembre 2007.

(Boite ,**1987**) Boite R., Kunt M.," Traité d'électricité : Traitement de la Parole", Presses Polytechniques Romandes, Lausanne, Suisse, 1987.

(Brutti ,**2008**) Brutti A, Omologo M., Svaizer P., "Comparaison between different sound source localization techniques based on a real data collection", IEEE HSCMA 2008.

(Hoang ,**2009**) Hoang Tran Huy Do., "Real-time SRP-PHAT source location: implementations on a large-aperture microphone array", Master of science in the division of engineering, Brown university, September, 2009.

(Khennouf,**2010**) Khennouf S., "Système automatique pour l'orientation de caméra mobile vers des cibles sonores", Mémoire de Magister en électronique: contrôle des processus et robotique, USTHB, Alger, 2010.

(Kunt,**1996**) Kunt M., "Traitement numérique des signaux". Edition technique et documentation, Lavoisier, 3ème édition, Lausanne, Suisse, 1996.

(Larousse,**1969**) Larousse P., Dictionnaire encyclopédique : Nouveau petit LAROUSSE, Paris, 1969, pp.719- 956.

(Margueron,**2000**) Margueron Xavier, "Capteur pour la localisation d'une source sonore", Département Génie Electrique et Informatique Industrielle 1: Electrotechnique et Electronique de Puissance, INRIA Rhône-Alpes, 25 Avril-30 Juin 2000.

Les sites web :

(Matlab7:programmingfondamentals,**2011**)
"Matlab7: programmingfondamentals".
http//www.mathworks.com
(Consulté le 19/06/2011 12:53)

(La *localisation* de la *source sonore*,**2012**) "La *localisation* de la *source sonore*"
http://cours.musique.umontreal.ca/MUS3321/MUS3321_SLIDES/MUS3321cours9_localisation.pdf
(Consulté le28/02/2012 14:36)

(*Localisation* d'un point d'émission d'une *source sonore* dans l'espace,**2012**) "*Localisation* d'un point d'émission d'une *source sonore* dans l'espace"
http //www.phaz.mc/edu/acoustique/localisation.html
(Consulté le22/02/2012 22:42)

(Pourquoi est-ce plus dur de localiser les sons aigus que les sons graves?,2012) "Pourquoi est-ce plus dur de localiser les sons aigus que les sons graves? "
http://philipe.boeuf.pagesperso -organe.fr/robert/physique/localisation-son.htm
(Consulté le 22/02/2012 22:36)

(Correlation,wikipédéa,**2012**) http://fr.wikipédia .org /wiki/ cor % c3 % a9latio_(statistiques).
(Consulté le 05/04/2012 14:18)

(Autocorrelation,wikipédéa,**2012**) http://fr.wikipédia .org /wiki/ autocorr % c3 % a9latio # traitement-du-signal.
(Consulté le 05/04/2012 14:23)