الجمهورية الجزائرية الديمقراطية الشعبية

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE DR. YAHIA FARES DE MEDEA

FACULTE DE LA TECHNOLOGIE

Département du Génie Électrique



Mémoire de Fin d'Études

Présenté pour l'obtention du Diplôme
De Master en Télécommunications

Option : Systèmes Intelligents et Communicants

THEME

La Séparation Aveugle des Signaux non-stationnaire en utilisant un banc de filtre Gammatone

Proposé et dirigé par,
Mr. ESSAID Billel

Préparé et présenté par,
Melle. Faïza BENMOULOUD
Melle. Meriem EULDJI

Année Universitaire 2016 – 2017

Remerciement

Avant tout, nous remercions le Dieu tout puissant de nous avoir donné la force et la volonté pour achever ce modeste travail.

Nous tenons à remercier particulièrement nos parents notre succès demeure de loin.

Nous nous devons remercier sincèrement notre promoteur Monsieur Essaid Billel pour son suivi, ses orientations et ses précieux conseils et qui a su nous faire profiter de sa grande expérience.

Nous tenons à remercier les membres de jury d'avoir accepter d'examiner ce modeste travail.

Nous n'oublions pas de remercier tous les enseignants qui ont suivi notre éducation pendant toute notre vie éducative.

Notre dernière remerciement s'adressent à toute personne ayant contribué, de prés ou de loin, a la réalisation de ce travail.

Faïza et Meriem

Dédicace

Je dédie ce modeste travail:

A mes très chers parents qui veille pour me voir réussir, et qui n'ont jamais cessé de m'encourager, de me conseiller …

Ils m'ont tout donné. Puisse Dieu, le tout puissant, vous préserve et vous accorder santé, longue vie et bonheur.



A mon cher frère : Mouaouia

sans oublié ma grand-mère que j'aime.

« Que Dieu me protège ma petite chère famille »

A tous les membres de la famille Benmouloud et Bailiche, petits et grands. En souvenir de toutes les joies et forces qui unissent notre chère famille

A ma très chère partenaire et amie Meriem.

A ma très chère amie Assia et toutes mes collègues Meriem, Chehera, Hadjer …

A toute personne qui de près ou de loin, a participé à ma Formation.

Aux membres de promotion.

Faïza

Dédicace

Je dédie modeste travail a tous ceux qui me portent dans leurs cœurs et m'accompagnant dans leurs pensées.

A ceux dont personne ne peut remplacer les sacrifies qu'ils ont consentis pour mon éducation et mon bien être : mes très chers parents.

A mon cher frère : Zaki , et mes chères sœurs Manel et Hadjer.

A tous mes oncles, tantes, cousins et cousines, en souvenir de toutes les joies et forces qui unissent notre chère famille «euldji et moulai mostepha ».

A tous mes amis, pour votre sincère amitié, votre soutien permanent me remonte le moral et vos conseils m'incitent à relever les défis.

A ma très chère partenaire et amie faiza.

A mon cher ami Abdou.

A mes très chers amis Nasoum, Lamia, Nabila, Nassima, Ommaima, et toutes mes collègues...

A toute personne qui de près ou de loin, a participe a ma formation.

Meriem

ملخص:

نقدم في هذه المذكرة، طريقة جديدة لفصل مصادر الصوت باستخدام خوارزمية MGD IS-NMF2D.

يستند تطوير خوارزمية الفصل على تمثيل مجال زمني – ترددي باستخدام مرشحات غاما. و بالخصوص نبين أن الإشارة السمعية ذات مستقبل واحد - ممزوج أكثر قابلية للفصل في مجال زمني – ترددي غير منتظم. كما سيتم عرض، بعض نتائج المحاكاة لتقييم أداء هذه الطريقة.

الكلمات المفتاحية: مرشحات غاما، فصل المصادر، الميل التنازلي.

Résumé:

Nous présentons dans ce mémoire, une nouvelle méthode de séparation de sources audio en utilisant l'algorithme MGD IS-NMF2D.

Le développement de cet algorithme de séparation est basé sur une représentation Temps-Fréquence (TF) en utilisant le banc de filtres Gammatone. En particulier, nous montrons que le signal audio mono-mixte est considérablement plus séparable dans le domaine TF non uniforme. Nous présenterons également, quelques résultats de simulation qui permettant d'évaluer les performances de cette méthode.

Mots clés : filtre Gammatone, Séparation de source, gradient de descente.

Abstract:

We present in this work, a new method for separation of audio sources using algorithm MGD IS-NMF2D.

The development of this separation algorithm is based on a Time-Frequency (TF) representation using the Gammatone filter bank. In particular, we show that the mono-mixed audio signal is considerably more separable in the non-uniform TF domain. We will also present some results of simulation which evaluating the performance of this method.

Key words: Gammatone filter, sources separation, decent gradient.

Table de matières

Liste des figures…….....III

Liste des tableaux……...IV

Liste des Sigles et abréviations……..……V

INTRODUCTION GENERALE:.....1

Chapitre I : Généralités sur la séparation des sources Aveugle.

I.1. Introduction :……3

I.2. Objectifs et Historique :……3

I.3. Problème de séparation de sources :……5

I.4. La nature du mélange :……6

I.4.1. Mélange instantané :……..7

I.4.2. Modèle du mélange Convolutif :……...8

I.4.3. Mélange anéchoïque :.....8

I.5. La non stationnarité :……9

I.6. Algorithmes de séparation aveugle de sources :……...9

I.6.1. L'algorithme SOBI :…… 9

I.6.2. L'algorithme COM2 :……..…10

I.6.3. L'algorithme COM1 :……..11

I.6.4. L'algorithme JADE :……11

I.7. Conclusion :……12

Chapitre II : Le banc de filtre Gammatone.

II .1. Introduction :.....13

II .2. Description du système auditif humain : 13

II .2.1. L'Oreille Externe : 13

II .2.2. L'Oreille Moyenne :.....14

II .2.3. L'Oreille Interne : 14

II .3. La cochlée:……..……...14

II .4. Filtres auditifs : 16

II .5. Masquage auditif :.....20

II.6. Le banc de filtres Gammatone:.....21

II .6.2. Le filtre Gammatone dans le domaine temporel : 22
II .6.3. Le filtre Gammatone dans le domaine fréquentiel : 23
II .6.4. Largeur de bande rectangulaire équivalente du filtre Gammatone : 24
II .6.5. Compensation de phase : 25
II .7. Conclusion: 26

Chapitre III : La simulation de la Séparation de Source Aveugle.
III.1. INTRODUCTION : 27
III.2. Qu'est-ce que MATLAB ? 27
III.2.1. Le son sous Matlab : 28
III.2.2. Lecture d'un fichier son : 28
III.2.3. Ecriture d'un fichier son: 28
III.3. Les données de traitement : 28
III.4. Banc de filtres Gammatone : 29
III.5. Choix de l'échelle fréquentielle : 29
III.6. Résultats : 30
III.7. Transformée de Fourier discrète : 31
III.8. Le cochléogramme : 32
III.9. Algorithme de séparation de sources : 33
III.10. Estimation des sources : 34
III.11. Résultats de simulation : 35
III.12. Conclusion : 36
CONCLUSION GENERALE : 37
Bibliographie : 38

Liste des figures

Chapitre I : Généralités sur la séparation des sources Aveugle.

Figure I. 1. la séparation de source……3

Figure I. 2. La problématique de la séparation de sources…….6

Figure I. 3. mélange linéaire instantané…….7

Figure I. 4. Problème multiple concernant le mélange convolutif…….8

Chapitre II : Le banc de filtre Gammatone.

Figure II. 1. Le système auditif.....13

Figure II. 2. Anatomie de l'oreille humaine. 14

Figure II. 3. Carte tonotopique de la cochlée. L'échelle de couleurs représente l'intensité du mouvement de la membrane basilaire pour une excitation dans la bande des 3 kHz. 15

Figure II. 4. Les mécanismes de perception humaine. 16

Figure II. 5. Modèle de seuil de détection pour un neurone du nerf auditif. Le seuil à la fréquence cible est de 20 dB SPL. D'après Hartmann. 17

Figure II. 6. Pattern d'excitation de deux sons purs de niveau modéré. F1 = 500Hz (50 dB SPL) et F2 = 1200Hz (20 dB SPL).....18

Figure II. 7. Pattern d'excitation de deux sons purs de fort niveau. F1 = 500Hz (90 dB SPL) et F2 = 1200Hz (60 dB SPL).....19

Figure II. 8. Réponse en fréquence des fonction G(z) (en bleu) et K(z) (en rouge) pour ã=0.9693 + 0.2104i, correspondant à une fréquence centrale de 1,5kHz et un largeur de bande de 50Hz.....22

Chapitre III : La simulation de la séparation des sources Aveugle.

Figure III. 1. Le tracé du mélange de deux signaux à séparés.....29

Figure III. 2. *Réponse en fréquence d'un banc de filtre Gammatone 12 bandes, f* min *=70Hz.* 30

Figure III. 3. *Réponse implusionelle de la bande passante N°3. Les paramètres sont* 31

Figure III. 4. Le *Signal de filtre gammatone filtré de la bande passante N°3.* 32

Figure III. 5. La sortie de l'énergie *d'un filtre Gammatone de la bande passante N°3.* 33

Figure III. 6. La première Source estimée.....36

Figure III. 7. La deuxième Source estimée.....36

Liste des tableaux

Tableau II .1: les données pour le filtre Gammatone……*25*

Liste des Sigles et abréviations

- $|\bullet|$ *:* module ou Valeur absolue.
- $\sum$: Sommation
- A : matrice de mélange instantané composée de coefficients scalaires réels aij.
- $Cumr(.)$: cumulant d'ordre r.
- H *:* entropie.
- λ *:* valeurs propres d'une matrice.
- $x(t)$ resp. $x(n)$: vecteur colonne composé de signaux d'observation $xi(t)$ resp. $xi(n)$.
- T : nombre d'échantillons de chaque observation.
- t : variable du temps.

Abréviations

TFD : Transformée de Fourier Discrète.

COM2 : COntrast Maximization2.

ERB : Largeur de bande rectangulaire équivalente.

IIR : Réponse impulsionnelle infinie.

JADE : Joint Approximate Diagonalization of Eigen-matrices.

MGD IS-NMF2D : multiplicative Gradient Descent Itkura Saito-Nonnegative Matrix Factorization 2dimensional.

SOBI : Second Order Blind Identification.

TF: Temps Fréquence.

TFD : Transformée de Fourier Discrète.

INTRODUCTION GENERALE:

Le traitement de la parole représente un élément très important dans les sciences de l'ingénieur. Il se situe entre le traitement du langage (c'est-à-dire du traitement de données symboliques) et le traitement du signal numérique. Cette discipline scientifique a connu depuis les années 60 une expansion fulgurante, liée au développement des moyens et des techniques de télécommunications. Comme la parole représente un vecteur d'information dans notre société humaine, elle occupe une place privilégiée, d'où l'importance particulière du traitement de la parole.

Le traitement de signal a pour vocation d'extraire l'information véhiculée par le signal. Ceci s'appuie sur la connaissance soit du signal, soit du système, soit des deux. Les systèmes de traitement audio voient en leur entrée, selon le cas, un ou plusieurs signaux. Cependant, l'environnement externe impose à ces derniers, une transformation au cours de leur propagation telle qu'une superposition avec d'autres signaux de l'environnement. Un cas typique de tels systèmes se trouve dans les organes auditifs de l'être humain. En effet, l'être humain est capable de focaliser, dans le mélange provenant de son environnement, sur une des sources de signal qu'il reçoit à son oreille. Dans cet objectif, plusieurs travaux ont été menés dans ce domaine et plus précisément avec la terminologie de séparation aveugle de sources : "Blind Source Separation : BSS".

Au cours des dernières décennies, beaucoup d'attention a été accordée à la séparation aveugle de sources. Dans plusieurs situations, il est souhaitable de récupérer toutes les sources à partir des mélanges enregistrés, ou du moins de séparer une source particulière.

La séparation de sources est un sujet de recherche très actif en raison de son intérêt théorique et de ses nombreuses applications pratiques. Dans la vie courante, nous pouvons citer bon nombre d'applications de séparation de sources correspondant à différents types de signaux rencontrés (signaux de paroles, signaux biomédicaux, etc.). A titre d'exemple :

- Dans le domaine *médical*, elle est utilisée pour l'extraction non invasive de l'Electro Cardio-Gramme (ECG), l'Electro Myo-Gramme (EMG) ou l'Electro Encéphalo-Gramme (EEG).

- Dans le domaine de la *communication*, le problème le plus connu (en acoustique) est certainement "l'effet cocktail party". Lors d'une fête, il s'agit de séparer la parole de notre interlocuteur parmi les nombreuses voix, la musique et autres bruits de fond.
- On trouve aussi des applications en téléphonie mobile où la parole de notre interlocuteur peut être issue d'un environnement bruyant (musiques, voitures qui passent dans la rue, etc...). Le signal enregistré est parfois considérablement "abimé". Il convient alors de séparer la source utile (parole) de la source perturbante (bruit de fond) pour parvenir ensuite à atténuer ces bruits de fond.
- Dans le domaine de la *sismologie*, on retrouve également le problème de séparation de sources dans le sens où le signal émis, par une vibration ou une explosion, est inconnu et observé au travers d'un mélange reflétant des sources multiples. La séparation de sources doit permettre aux sismologues, à partir d'enregistrements et moyennant des renseignements sur la fonction de transfert du sous-sol, d'obtenir des informations sur la source et de la modéliser.

Cette liste d'applications est non exhaustive. Il y en a bien d'autres, comme en astrophysique, pour le contrôle du trafic aérien.[1],[2]

Notre mémoire est organisé comme suite :

- Dans le premier chapitre, nous présenterons le problème de séparation de sources en expliquant son principe, et en donnant la nature de mélanges des sources, ainsi les algorithmes les plus connus dans ce domaine.
- Dans le deuxième chapitre, nous présentons le système auditif humain, ainsi que le banc de filtres Gammatone qui est basé sur le principe de l'oreille, ce dernier banc de filtres est à la base de l'algorithme utilisé dans notre mémoire pour séparer les sources.
- Dans le dernier chapitre, nous présentons l'algorithme de séparation de sources avec les résultats de simulations obtenus

Nous terminons notre mémoire par une conclusion générale.

I.1. Introduction :

La séparation aveugle de sources est un problème qui consiste à retrouver des signaux a partir de leurs mélanges les solutions proposées pour ce problème varient suivant la nature des sources, le type de mélange (instantané ou convolutif) et le rapport entre le nombre de capteurs et de sources.

Dans ce chapitre, nous nous discuterons sur le principe de la séparation aveugle de sources, ainsi que les différents algorithmes existants dans ce domaine.

I.2. Objectifs et Historique :

Le problème de séparation de sources consiste à recouvrer des signaux utiles provenant de différentes sources, qui ont été mélangées ; c'est-à-dire, filtrées et additionnées en se propageant vers un ensemble de capteurs. Ce problème se rencontre dans plusieurs applications.

L'approche aveugle de la séparation, offre en outre l'avantage de ne requérir que des hypothèses faibles sur le mélange : en dehors de sa structure générale, souvent supposée linéaire, aucun paramètre n'est supposé connu apriori.

La figure I.1 donne une idée générale sur le principe de la séparation de sources.

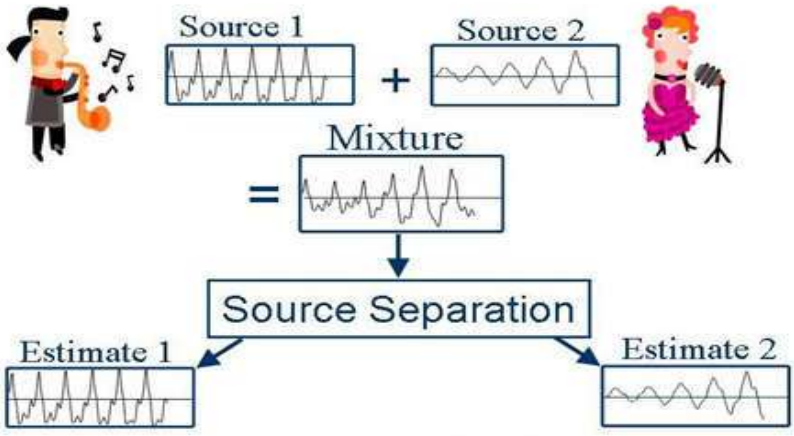


Figure I. 1. *La séparation de source.*

La séparation aveugle de sources a été formalisée pour la première fois par des chercheurs français en traitement de signal au milieu des années 80. Depuis lors, beaucoup de chercheurs dans le monde entier se sont intéressés à ce problème, vu la diversité de ses domaines d'application. Les recherches dans ce domaine ont été initiées

en 1985 par C. Jutten et J. Hérault. Ces derniers ont développé, à partir d'un problème de décodage du mouvement en neurobiologie, un algorithme de séparation à base d'un réseau neuronale. Cet algorithme cherche à résoudre le problème du mélange instantané sans bruit avec une matrice de mélange carrée et des signaux à valeurs réelles.

Depuis, une multitude d'algorithmes a été proposée. Ils peuvent être classés selon différents critères : mélange instantané ou mélange convolutif, linéaire ou non linéaire, statistiques d'ordre deux ou d'ordre supérieur.

En 1988 J. L. Lacoume et P. Ruiz ont proposé un algorithme minimisant une fonction objective basée sur les cumulants croisés des sorties d'un filtre inverse, qui sont les outils naturels de mesure d'indépendance.

En 1989, J. F. Cardoso et Souloumiac proposent de considérer les propriétés algébriques des cumulants d'ordre quatre en utilisant l'information d'ordre deux pour rendre unitaire la matrice de mélange puis de l'identifier par des techniques de diagonalisation. En revanche, ils montrent que les cumulants d'ordre quatre sont suffisants pour permettre, à eux seuls, d'estimer la matrice de mélange.

L'approche développée par P. Comon, en 1990, élargit la théorie de l'analyse en composantes principales (en anglais : PCA, Principale Components Analysis) à l'analyse en composantes indépendantes et introduit la notion de fonction de contraste. L'auteur propose comme contraste à maximiser la somme des cumulants d'ordre quatre. Comon montre par ailleurs, qu'il suffit de reconstruire des sources indépendantes deux à deux pour estimer, à une matrice triviale près, des sources mutuellement indépendantes.

Indépendamment des travaux de P. Comon, M. Gaeta et J. L. Lacoume proposent de résoudre le problème de séparation par une technique du maximum de vraisemblance approchée.

B. Laheld et J. F. Cardoso ont développé en 1993 une famille d'algorithmes adaptatifs basés sur la notion de gradient matriciel relatif toujours en 1993, survient la méthode JADE (Joint Approximate Diagonalization of Eigen-matrices), au travers de la quelle Cardoso et Souloumiac présentent une solution algébrique à la maximisation de leur contraste basé sur les cumulants d'ordre 4. Ils étendent par la même occasion l'algorithme de Jacobi dans le but de diagonaliser conjointement un ensemble de matrices.

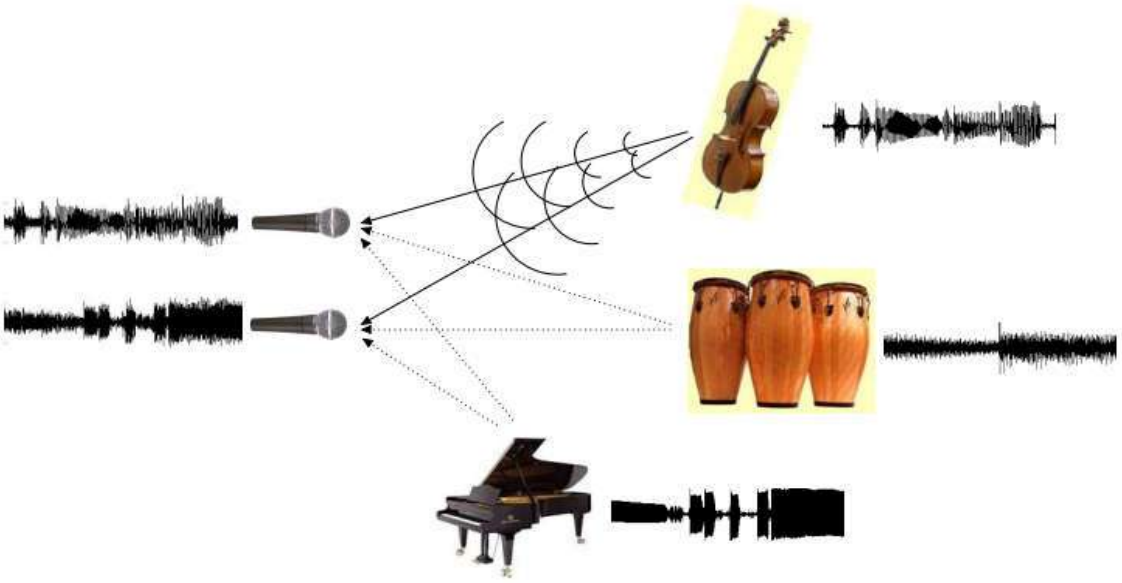
En 1995, Delfosse et Loubaton introduisent le concept de déflation en proposant une manière originale d'optimiser le contraste uni-source d'ordre 4 sur lequel s'appuie

leur méthode. Deux années plus tard, Hyvärinen et Oja présentent l'algorithme Fast-ICA. La version initiale de cet algorithme exploite pleinement le principe de déflation et la mesure de non-gaussianité, plus connue sous le nom denéguentropie, approximée par la valeur absolue du kurtosis de la source estimée.

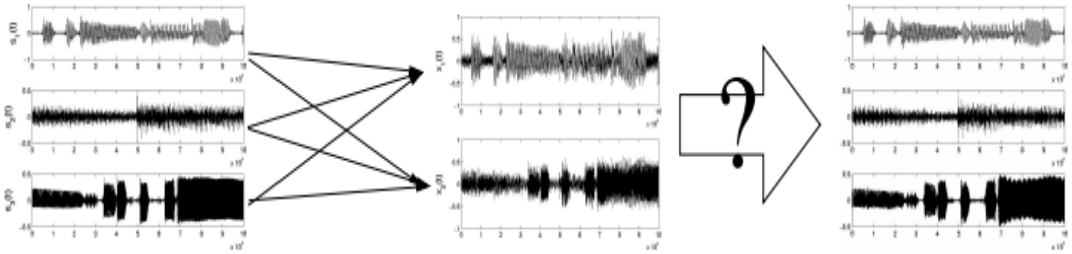
Un certain nombre d'auteurs parmi eux K. Kokkinakis en 2006, ont utilisé des critères en combinant le domaine temporel et domaine fréquentiel avec la mise en œuvre de la rapidité des calculs. Un des moyens les plus efficaces de la séparation aveugle de sources dans le cas des mélanges convolutifs utilise les relations statistiques de la densité de puissance du signal, au lieu de résoudre la séparation de sources dans chaque bande de fréquence pour le cas du mélange instantané. Kim (2006) propose une version d'algorithme multidimensionnel pour l'estimation de la densité (2007) et en exploitant les dépendances en fréquence d'ordre supérieur [1].

I.3. Problème de séparation de sources :

Dans ce mémoire, nous nous intéressons à l'analyse de scènes sonores, et en particulier à la séparation aveugle de sources appliquée aux signaux audio. Nous considérons la configuration suivante : une scène est composée de sources sonores et observée par l'intermédiaire de capteurs, typiquement des microphones. Le signal observé est un mélange, résultat de la propagation et de l'addition des ondes de chacune des sources jusqu'aux capteurs. Le problème de la séparation de sources consiste à extraire chacun des signaux sources à partir de l'observation du mélange de ces sources. La *Figure I.2* illustre la problématique de la séparation de sources dans le cas d'un mélange de trois sources captées par deux microphones.



(a) scène sonore composée de 3 sources captées à l'aide de 2 microphones.



(b) *La séparation des 3 sources à partir du mélange sur 2 canaux*

Figure I. 2. *La problématique de la séparation de sources.*

Il existe un grand nombre de méthodes de séparation de sources, adaptées à des problématiques déférentes. Ces problématiques se caractérisent par des hypothèses sur le mélange et sur les sources qui varient suivant l'application et le degré de réalisme que l'on souhaite atteindre [3].

I.4. La nature du mélange :

La nature du mélange représente les hypothèses que l'on fait sur l'environnement acoustique dans le quel l'enregistrement est effectué. En particulier, les enregistrements qui ont lieu dans un milieu naturel sont affectés par de multiples échos, qui ont pour conséquence d'introduire des directions multiples pour chacune des sources. Si l'enregistrement est

effectué dans une salle anéchoïque, où l'on peut supposer qu'il n'y a plus d'écho, par contre les distances d'une source à chaque microphone étant différentes, les signaux des sources n'arrivent pas au même instant aux différents microphones. On définit trois hypothèses sur le mélange qui correspondent aux trois scénarios précédents : mélange instantané, mélange convolutif et mélange convolutif anéchoïque [3].

I.4.1. Mélange instantané :

C'est le cas le plus simple dans lequel les sources sont uniquement caractérisées par une différence de niveau d'intensité entre les canaux.

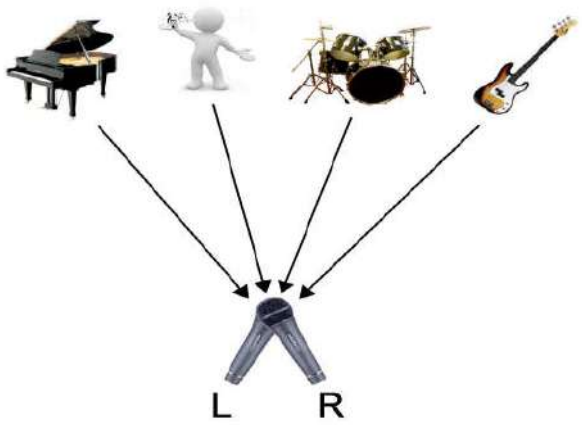


Figure I. 3. *mélange linéaire instantané.*

La séparation de sources est une technique de traitement des signaux (ou images) multi-capteurs, dans la quelle on postule un modèle où une séquence d'observations M-dimensionnelles $x(t),\ t = 1, \dots, T$ est modélisée par :

$$x(t) = A\, s(t) + w(t) \qquad (I.1)$$

Où $x(t)$ représente les M signaux observés $x(t) \triangleq [x_1(t), \dots, x_M(t)]^T$

Le signal $x(t)$ correspond au mélange de N signaux sources $s(t) \triangleq [s_1(t), \dots, s_N(t)]^T$ à travers une matrice de mélange $A \triangleq [a_1, \dots, a_N]$[1], [4].

I.4.2. Modèle du mélange Convolutif :

L'équation (*I.1*) décrit un mélange instantané. Malheureusement ce modèle est plutôt incomplet dans le cas de sources enregistrées dans un environnement acoustique

réel, où les sources sont mélangées convolutivement. En ignorant le bruit *w(t)*, le problème peut être formulé comme suit:

$$x(t) = \sum_{k=0}^{K} H(k).s(t-k) + w(t) \qquad (I.2)$$

Où les *H(k)* sont des matrices de taille $M \times N$ pour $k \in [0;K]$ qui représentent les coefficients de la réponse impulsionnelle du canal.

La matrice polynomiale $H(k) = \sum_{k=0}^{K} z^{-k}$ est supposée irréductible [4].

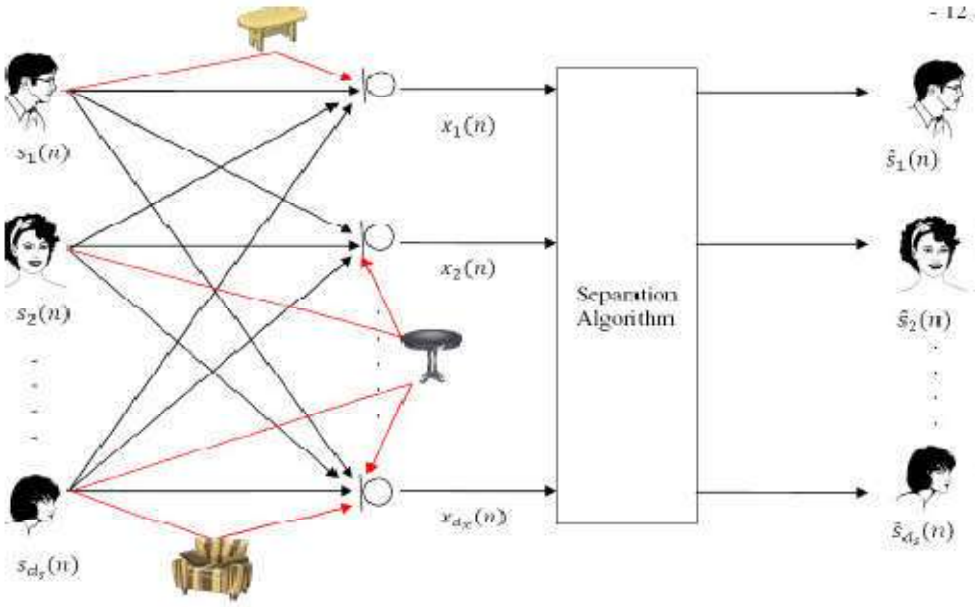


Figure I. 4. *Problème multiple concernant le mélange convolutif.*

I.4.3. Mélange anéchoïque :

C'est une généralisation du cas instantané, dans lequel on considère que les sources arrivent à chaque capteur à des instants différents, si bien que chaque source est caractérisée par un délai entre les canaux.

Le cas anéchoïque peut être considéré comme un cas de mélange convolutif, dans lequel on considère des chemins multiples entre chaque source et chaque capteur [1].

I.5. La non stationnarité :

La résolution du problème de séparation de sources dans le cas du mélange convolutif s'effectue souvent par des approches qui transforment le mélange convolutif en un ensemble de mélanges instantanés. Ces approches transposent le problème dans le

domaine de Fourier par une TFD. En conséquence, il est possible d'utiliser les méthodes des mélanges instantanés à chaque canal de fréquence puis de reconstruire les sources en ayant l'ensemble des solutions fréquentielles.

Le problème semble, ainsi, simplifié. Cependant, la mise en œuvre de cette approche se heurte à certains problèmes.

En effet, en plus du problème d'ambiguïté d'échelle et de permutation intrinsèque à la séparation de sources, la plupart des méthodes de séparation reposent sur l'hypothèse de non gaussianité des sources. Or, dans le cas général, il est bien connu que la TFD produit des signaux qui sont asymptotiquement gaussiens. La distance à la gaussianité des signaux, après transformée de Fourier et a proposé un test de mesure pour la validation d'applicabilité des algorithmes dans le domaine fréquentiel. Par ailleurs, Pham et Cardoso ont proposé une alternative qui permet de réaliser la séparation de sources tout en ignorant leur non gaussianité.

Cette méthode exploite la nonstationnarité des sources, comme dans le cas des sources audio par exemple, pour construire un ensemble de statistiques différentes permettant d'accomplir la séparation de sources qu'elles soient gaussiennes ou non [5].

I.6. Algorithmes de séparation aveugle de sources :

I.6.1. L'algorithme SOBI :

L'algorithme SOBI exploite plusieurs matrices de covariance associées à différents retards non nuls après blanchiment des observations. En outre, les sources doivent être temporellement corrélées, posséder des spectres différents et le bruit doit être temporellement blanc. La matrice de covariance des observations blanchies$\{x[m]\}$, $m \in \mathbb{N}$est alors donnée pour un retard $\tau_j \in \mathbb{N}^*$ par :

$$R_z[\tau_j] = E\left[z[m]z[m-\tau_j]^T\right] - E[z[m]]E\left[[m-\tau_j]^T\right] = \hat{H}R_s[\tau_j]\hat{H}^T \qquad (I.3)$$

Où $\hat{H}$ représente la matrice de mélange blanchie orthogonale, et $R_s[\tau_j]$ désigne la matrice de covariance des sources associée au retard τ_j, qui a la propriété d'être diagonale.

I.6.2. L'algorithme COM2 :

La première classe de méthodes est basée sur la technique de l'Analyse en Composantes Indépendantes, utilisant des outils d'ordre supérieur pour construire des

fonctions de contraste, les quelles exploitent les propriétés des cumulant des sources supposées indépendantes. C'est P.Comon qui a étendu la notion de contraste utilisée auparavant dans les systèmes de déconvolution aveugle au cas des vecteurs aléatoires.

Il a proposé dans une méthode baptisée COM2 (COntrast Maximization2) consistant à maximiser une fonction de contraste notée $\psi_R^2(y)$ (Il s'agit de l'une des premières fonctions de contraste), et définie comme la somme des modules au carré des cumulants d'ordre $R \geq 3$ des sources estimées $y(t)$:

$$\psi_R^2(y) = \sum_{i=1}^{n} |C_R\{y_i\}|^2 \qquad (I.4)$$

Où n est le nombre de sources et $C_R\{.\}$ représente le cumulant d'ordre R d'un signal stationnaire. La maximisation de la fonction contraste $\psi_R^2(y)$ a conduit à l'algorithme ICA.

De même que SOBI, COM2 nécessite un blanchiment des observations. Puis il faut maximiser le contraste. Une optimisation de ce contraste repose sur le théorème suivant:

Soit s un vecteur aléatoire de composantes indépendantes, parmi les quelles au plus une est gaussienne, et dont les densités de probabilité ne sont pas réduites à une masse ponctuelle. Soit T une matrice orthogonale de taille $(P \times P)$etc. le vecteur défini par $z=T.s.$ Alors, les trois propriétés suivantes sont équivalentes :

- les composantes de z sont indépendantes deux à deux,
- les composantes de z sont mutuellement indépendantes,
- T est une matrice triviale.

Ce théorème nous dit qu'au vu des indéterminations intrinsèques au problème de la Séparation Aveugle de Sources, il suffit juste de chercher à reconstruire des sources indépendantes deux à deux.

Une autre façon d'introduire les cumulants marginaux en tant que critère d'optimisation consiste à remarquer que les variables aléatoires sont indépendantes si tous Leurs cumulants croisés sont nuls. On peut donc être tenté de minimiser l'ensemble des cumulants croisés à un ordre donné.

I.6.3. L'algorithme COM1 :

Après un blanchiment des observations le contraste suivant est optimisé :

$$\psi_R^2(y) = \lambda \sum_{i=1}^{n} |C_R\{y_i\}| \quad , o\grave{u}\ \lambda = \pm 1 \qquad (I.5)$$

La solution itérative proposée pour COM2 peut être utilisée afin d'optimiser le contraste précédent. Cependant, une solution algébrique a été proposée. Notons toute fois que COM1 nécessite que les sources aient des kurtosis de même signe.

L'idée de l'algorithme COM1(α1) ou COM2(α2) dans $\psi_R^2(y) = \sum_{i=1}^{n} |C_R\{y_i\}|^{\alpha}$ est de remarquer que ce maximum peut être calculé entièrement analytiquement en dimension2, dans un certain nombre de cas. L'intérêt de l'algorithme COM1 réside dans une propriété établissant que la valeur absolue peut être supprimée lorsque les kurtosis des sources sont tous de même signe. C'est ce qui sera toujours implicitement supposé dans toutes les mises en œuvre de COM1.

I.6.4. L'algorithme JADE :

Il est basée sur l'utilisation de cumulants d'ordre quatre impliquant à la fois des auto-cumulants et des cumulants croisés :

$$J(y) = \sum_{i1,i2,i3=1}^{n} |cum\{y_{i1}, y_{i1}, y_{i2}, y_{i3}\}|^2 \qquad (I.6)$$

où l'opérateur *cum{.}* désigne un cumulant d'ordre quatre. La fonction de contraste précédente est appelée « contraste JADE ». Maximiser cette fonction revient à diagonaliser conjointement un ensemble de n^2 matrices.

L'algorithme JADE nécessite aussi un blanchiment des observations, puis la maximisation du contraste.

Une solution algébrique consiste à diagonaliser conjointement l'ensemble des matrices Propres M_p construites à partir des vecteurs propres associés aux P plus grandes valeurs propres de la matrice de quadri-covariance [1].

I.7. Conclusion :

Le concept de la séparation aveugle de source a été introduit au milieu des années 1980. Dans ce chapitre nous avons présenté la problématique de séparation aveugle de sources et plusieurs types de mélanges.

II .1.Introduction :

Dans ce chapitre nous allons présenter le principe de base des filtres Gammatone. Le banc de filtres Gammatone est un ensemble de filtres, chacun d'eux peut filtrer un signal dans une bande de fréquence appropriée, ces bande sont conçus pour imiter le système auditif humain.

II .2.Description du système auditif humain :

Le système auditif humain est assez complexe. En effet, il est sensible aux plus fins des mouvements aléatoires des molécules d'air en contact avec le tympan. De plus, il peut supporter des bruits très forts et peut extraire des faibles sons en présence de bruits énormes.

Il peut aussi distinguer et différencier plusieurs sons. Le système auditif comprend quatre parties : l'oreille externe, l'oreille moyenne, l'oreille interne et les vois nerveuses [8].

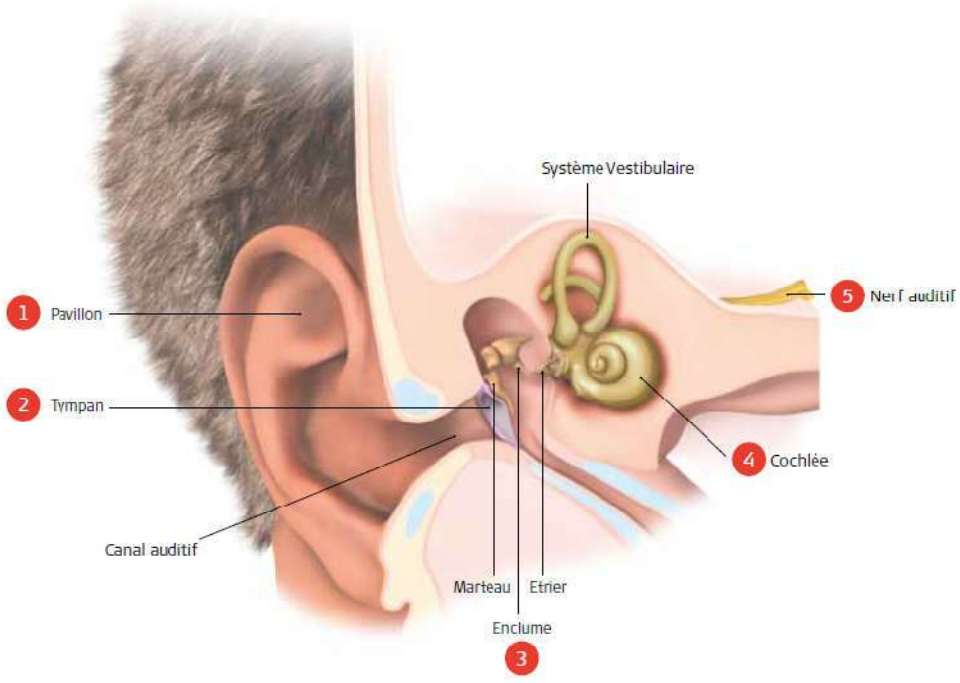


Figure II. 1. Le système auditif.

II .2.1.L'Oreille Externe :

L'oreille externe [E] composée par le Pavillon qui recueille et transmet les sons au canal auditif et le Canal Auditif (conduit auditif externe) qui dirige les sons dans l'oreille.

Elle amplifie les sons aigus entre 1.5 kHz et 7 kHz. Sa fonction est donc une préaccentuation du signal acoustique.

II .2.2. L'Oreille Moyenne :

L'oreille moyenne [M] assure l'adaptation d'impédance entre le milieu aérien et les liquides de l'oreille interne, elle composée par le Tympan (membrane tympanique) qui transforme les sons en vibrations et la Chaîne des osselets (marteau, enclume, étrier) transmet les vibrations à l'oreille interne.

II .2.3. L'Oreille Interne :

Oreille Interne [I] (cochlée) contient des fluides et des cellules très sensibles (cellules ciliées), avec des structures en forme de cils qui bougent grâce aux vibrations sonores. Le Système Vestibulaire contient des cellules qui contrôlent l'équilibre. Et le Nerf Auditif transmet l'information de la cochlée au cerveau.

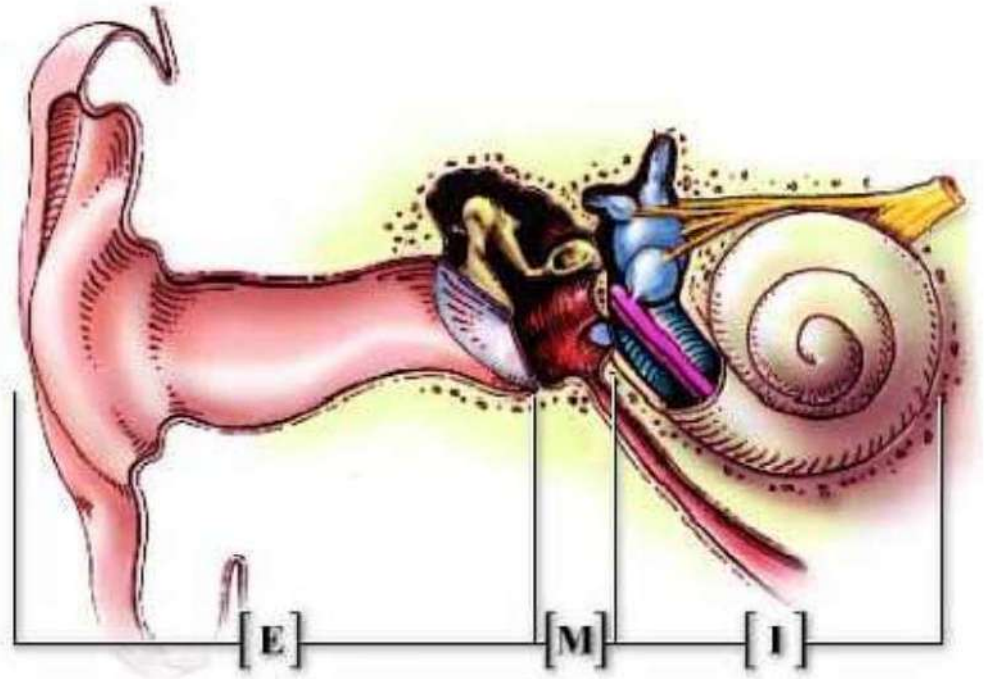


Figure II. 2. Anatomie de l'oreille humaine.

II .3. La cochlée :

C'est un tube qui s'enroule sur lui-même en deux tours et demi allant de la base à l'apex, comme une coquille d'escargot, figure II .3 Le tube a une longueur totale de 30 mm, une hauteur du milieu de la base au sommet de 3 à 4 mm, un diamètre du tube à l'origine de 2 mm et au sommet de 1 mm.

Ainsi, les vibrations mobilisent le tympan et la chaîne des osselets, qui transfèrent la vibration à la liquide périlymphe de la rampe vestibulaire. L'onde se propage de la base vers l'apex et retourne vers la base par la rampe tympanique. Les deux rampes entourent

le canal cochléaire, qui contient l'organe sensoriel auditif appelé organe de Corti, et qui baigne dans la liquide endolymphe. Le nerf auditif se ramifie pour se connecter aux cellules sensorielles de l'organe de Corti.

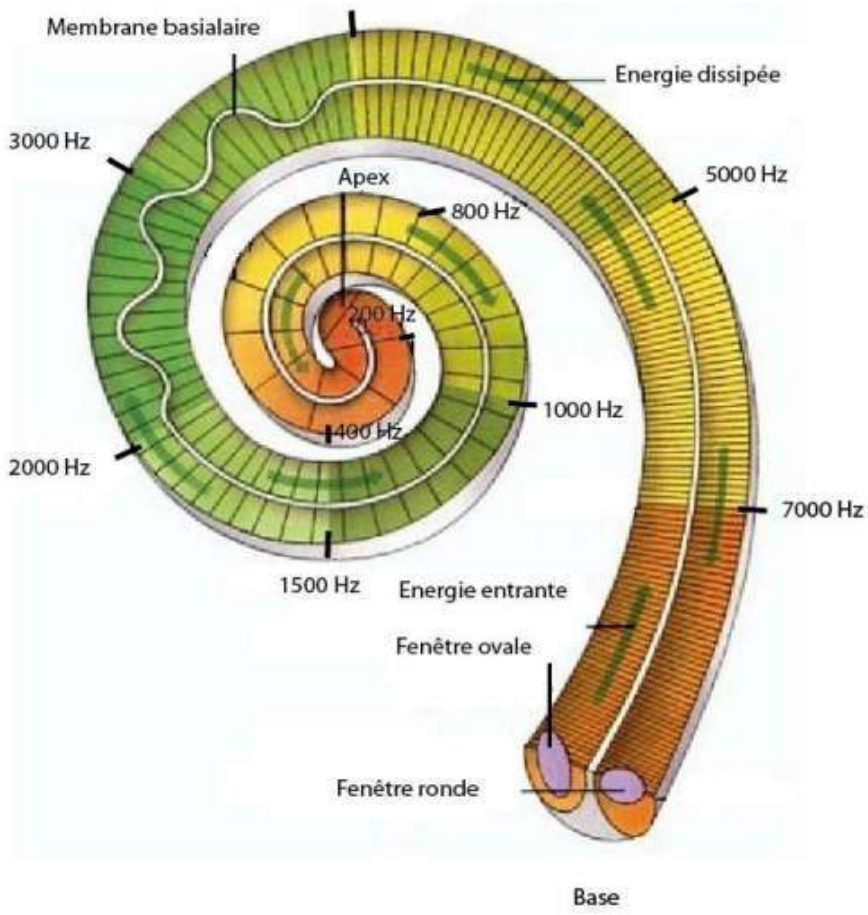


Figure II. 3. *Carte tonotopique de la cochlée. L'échelle de couleurs représente l'intensité du mouvement de la membrane basilaire pour une excitation dans la bande des 3 kHz.*

Les CCI disposées dans la cochlée, le long de la membrane basilaire, sont chacune sensible à une fréquence caractéristique différente. La tonotopie ainsi crée est l'aspect primordial que les différents modèles de cochlée cherchent à reproduire, chaque élément du banc de filtre modélisant la réponse de la membrane basilaire autour d'un point précis [8].

La figure suivante montre le mécanisme de perception auditif.

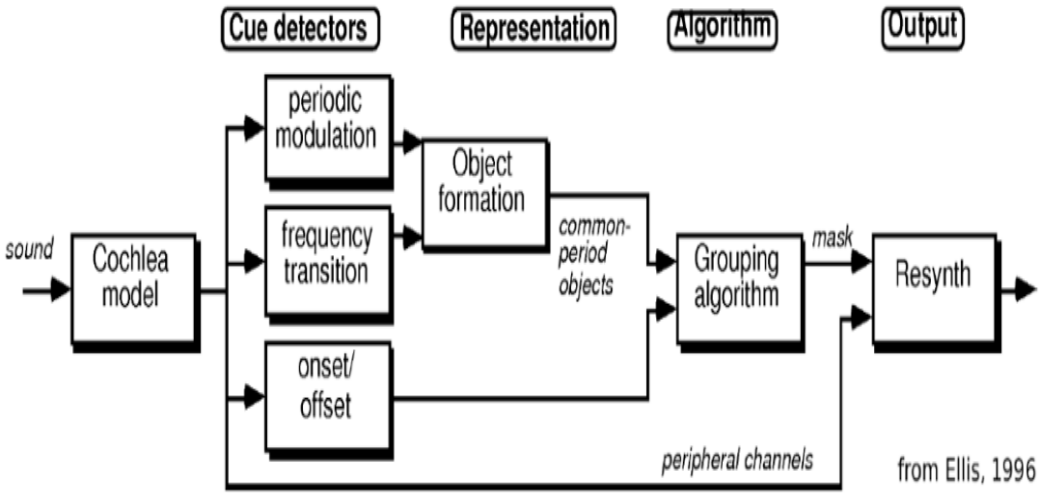


Figure II. 4. *Les mécanismes de perception humaine.*

II .4. Filtres auditifs :

La perception du son commence par la conversion d'un champ de pression acoustique en impulsions neuronales grâce aux cellules ciliées disposées le long de la membrane basilaire. Chacune de ces cellules correspond à une fréquence, et les caractéristiques du système sont très sélectives. A l'aide d'une fréquence pure, une expérience de neurophysiologie chez l'animal (où l'on enregistre l'activité d'une fibre du nerf auditif) permet de construire le seuil de détection pour un seul neurone : c'est la plus faible amplitude nécessaire pour exciter le nerf auditif en fonction de la fréquence. Une approximation est montrée figure II.5 On remarque un passe-bande pointu centré sur la fréquence d'excitation.

L'idée de la modélisation est donc de considérer la membrane basilaire et les cellules ciliées comme un banc de filtres passe-bande, appelés filtres auditifs. Dans le cadre de filtres linéaires, il est possible de retrouver le transfert en gain de ces filtres en inversant la courbe de seuil de détection.

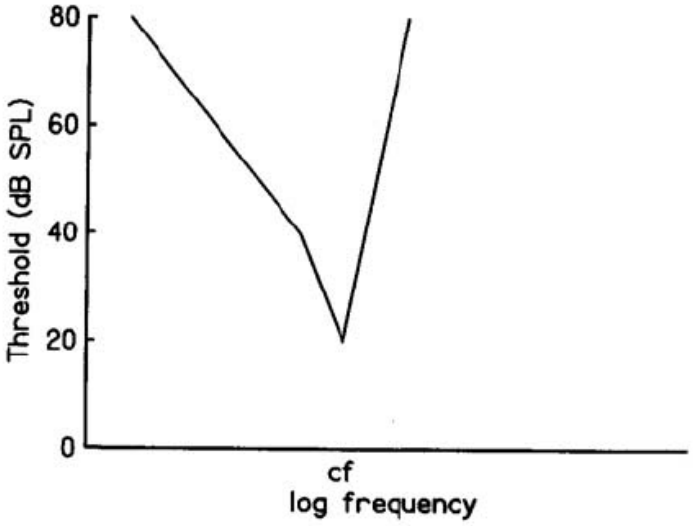


Figure II. 5. *Modèle de seuil de détection pour un neurone du nerf auditif. Le seuil à la fréquence cible est de 20 dB SPL. D'après Hartmann.*

Une telle approche est cohérente avec les phénomènes de masquage simultané. En effet, pour une fréquence d'excitation donnée, les fréquences proches sont sollicitées. Il est donc prévisible que le seuil d'audition sera ainsi augmenté pour un autre son pur aux abords de cette première fréquence, si celui-ci n'a pas assez d'amplitude, il est alors masqué.

Historiquement, les filtres auditifs ont été caractérisés en premier lieu par leur largeur de bande, sous l'hypothèse qu'ils pouvaient être modélisés par des filtres passe-bande idéaux, c'est-à-dire rectangulaires. Etant donné la complexité des phénomènes mis en jeu au niveau de la cochlée, cette hypothèse est évidemment très simplificatrice. En effet, des mesures plus précises des filtres auditifs ont montré leur caractère non linéaire et en particulier leur dépendance vis-à-vis du niveau des stimuli. Les filtres auditifs ont en réalité une bande passante plus étroite pour les niveaux modérés que pour les niveaux forts.

Cependant, dans ce cadre très complexe, une modélisation par des filtres passe-bande idéaux constitue une première approche pratique et permet de quantifier approximativement la résolution fréquentielle du système auditif. Moore et Glasberg, ont proposé une mesure de la largeur de ces filtres idéaux appelée *largeur de bande rectangulaire équivalente* (ou *Equivalent Rectangular Bandwidth* -ERB-). Pour chaque filtre auditif, la valeur d'ERB est définie comme la largeur d'un filtre passe bande idéal de même fréquence centrale qui, alimenté par un bruit blanc, produirait la même énergie en sortie. La mesure des ERB illustre la résolution fréquentielle du système auditif. La formule suivante donne la relation moyenne pour plusieurs normo entendant entre la fréquence centrale f et l'ERB en Hz :

$$ERB(f) = l + \frac{f}{q} \text{ ,avec } l = 24{,}7 \text{ et } q = 9{,}265 \qquad \text{(II .1)}$$

On peut définir une échelle des ERB. Chaque numéro d'ERB correspond alors à une distance fréquentielle d'une bande rectangulaire équivalente.

$$N°ERB(f) = q log_{10}\left(1 + \frac{f}{l.q}\right) \qquad \text{(II .2)}$$

augmente avec la fréquence, de telle sorte que cette largeur vaut approximativement 11% à17% de la valeur de la fréquence centrale. Ces caractéristiques influent sur la résolution fréquentielle du système auditif, en effet, la discrimination fréquentielle de l'oreille est meilleure pour les basses fréquences que pour les hautes.

Schématiquement, on peut résumer ainsi les propriétés générales des filtres auditifs:

- Le sommet de chaque filtre est étroit et à peu près symétrique en fréquence.
- Les bords sont larges et asymétriques, en particulier leur pente du côté des basses fréquences est moins forte.
- Plus le niveau du stimulus est élevé, plus les filtres sont larges.

La modélisation par les filtres auditifs permet également une représentation schématique de l'excitation de la membrane basilaire en réponse à un stimulus. Cette représentation est appelée *pattern d'excitation*, anglicisme tiré de l'expression *excitation pattern*. Moore et Glassberg ont proposé une méthode simple pour estimer le pattern d'excitation à partir de la sortie des filtres auditifs. Celui-ci est obtenu en relevant le niveau de sortie de chaque filtre en fonction de sa fréquence centrale.

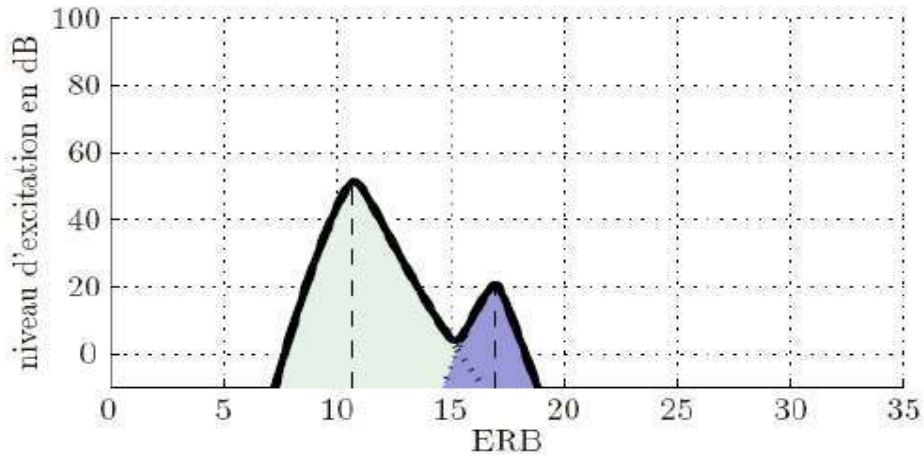


Figure II. 6. *Pattern d'excitation de deux sons purs de niveau modéré. F1 = 500Hz (50 dB SPL) et F2 = 1200Hz (20 dB SPL).*

La figure II .6 montre un exemple de différents patterns d'excitation :

- celui d'un son pur à 500Hz (10,7 en échelle ERB) et d'intensité 50 dB SPL indiqué par la courbe englobant la zone claire centrée à 10,7 ERB,
- celui d'un son pur à 1200Hz (17 en échelle ERB) et d'intensité 20 dB SPL indiqué par la courbe englobant la zone d'aire foncée centrée à 17 ERB,
- enfin le pattern d'excitation complexe composé des deux fréquences précédentes, représenté par la courbe supérieure en trait gras.

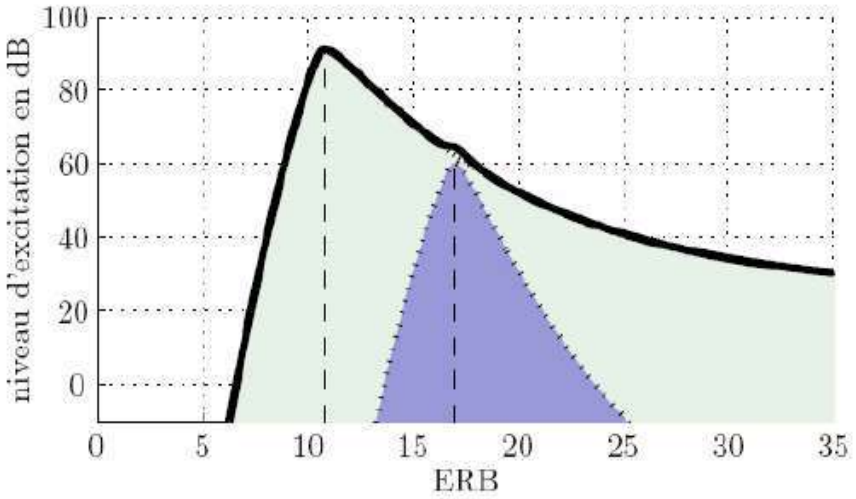


Figure II. 7. *Pattern d'excitation de deux sons purs de fort niveau. F1 = 500Hz (90 dB SPL) et F2 = 1200Hz (60 dB SPL).*

Le pattern d'excitation est un outil intéressant car il permet d'une part de retrouver la tonotopie de la membrane basilaire, et d'autre part d'illustrer avec clarté le phénomène de masquage auditif. En effet, le stimulus de la figure II.6présente deux pics fréquentiels distincts correspondant aux deux fréquences d'excitation. Si on considère maintenant le même exemple mais pour des sons purs rehaussés de 40dB chacun, on obtient le pattern d'excitation de la figure II .7. A ce niveau de stimulation, les filtres auditifs sont plus larges. On remarque alors que le pattern d'excitation du premier son pur indique une excitation beaucoup plus étendue sur la membrane basilaire que dans le premier exemple et surtout qu'en tout point de la membrane basilaire, ce premier apporte plus d'énergie que le second. Le second pattern d'excitation est ainsi complètement recouvert par le premier. Le pattern d'excitation résultant est pratiquement confondu avec le premier et n'indique plus qu'un seul pic fréquentiel : on voit donc ici clairement apparaître le phénomène de masquage.

Notons que l'effet d'élargissement des filtres avec le niveau d'excitation ne sera pas explicitement simulé dans le banc de filtre mais plutôt dans un étage de compression non linéaire qui sera placé après, et qui produit des effets qualitativement similaires [6].

II .5.Masquage auditif :

Le masquage auditif est un phénomène perceptif qui fait qu'un son pourtant audible n'est pas perçu, à cause de la présence d'un autre son : le signal non perçu est alors dit masqué. En termes psycho acoustiques, le son masqué est appelé son cible, et l'autre masqueur. De façon plus précise, le masqueur élève localement le seuil d'audition, le son cible nécessite alors plus d'énergie pour être perçu.

Il existe deux types de masquage : le masquage simultané ou fréquentiel, et le masquage temporel. Le masquage fréquentiel apparaît lorsque que les deux sons sont présentés simultanément : on observe une élévation du seuil d'audition pour les fréquences proches du masqueur. Le masquage temporel est par contre produit par la présence du masqueur juste avant ou juste après la cible : le seuil de détection est augmenté aux abords du masqueur. Le masquage fréquentiel est celui qui produit les effets de plus grandes amplitudes.

Le phénomène de masquage est à rapprocher du concept de 'bande critique auditive'.

Cette approche découle du traitement de l'information sonore dans la cochlée, et plus particulièrement de la vibration de la membrane basilaire. En effet, cet organe vibre en présence d'un son, et chaque point est sensible à une certaine fréquence pour laquelle l'amplitude de la vibration sera maximale. Ainsi, pour une fréquence donnée, la membrane basilaire est excitée a un point précis, entraînant une vibration de plus faible amplitude aux fréquences voisines : chaque point de la membrane basilaire peut alors être considéré comme un filtre passe bande d'une certaine largeur, la bande critique. Cette organisation, où chaque fréquence correspond à une localisation d'excitation, est appelée tonotopique.

Si la membrane basilaire est excitée à une certaine fréquence (par un masqueur), une nouvelle excitation à une fréquence proche (son cible) ne sera alors perçue que si elle sur monte la vibration induite par la première. Il n'est généralement admis que cette compétition entre excitations fourni la base des phénomènes de masquage.

La modélisation fonctionnelle du concept de bande critique se fera à l'aide d'un banc de filtre visant à reproduire l'excitation mécanique de la membrane basilaire. Le signal de chaque canal du banc de filtre correspond alors à la déformation de la membrane basilaire à

un endroit donné. A partir de l'observation des phénomènes de masquage, on va pouvoir déterminer la forme de ces filtres, leur largeur, et la façon de les disposer dans le banc. Il est important de remarquer que le modèle basé sur ces filtres ne sert pas seulement pour le masquage, puisque nous allons pouvoir calculer plusieurs attributs perceptifs qui ne sont pas directement liés au masquage. Le modèle simule en fait la variation de sélectivité fréquentielle du système auditif en fonction de la fréquence, ce qui est la principale différence des étapes précoces de l'encodage cochléaire avec une analyse de type transformée de Fourier à court terme [6].

II.6.Le banc de filtres Gammatone:

II.6.1. Introduction sur les filtres Gammatone :

Les filtres Gammatone sont souvent utilisés pour la modélisation du traitement du son dans la cochlée. Plusieurs implémentations de ces filtres sont connues, toutes proviennent de diverses approximations du filtre Gammatone original.

Pour notre projet, il a été retenu un banc de filtres Gammatone tout pôle d'ordre 4, à sortie complexe, proposée par Hohmann. Afin de pouvoir générer rapidement le filtre, ce dernier a donc été simplifié par omission du numérateur, en effet, il est dérivé de la version échantillonnée de la réponse impulsionnelle Gammatone analogique:

$$g_\gamma(n) = n^{\gamma-1}.\tilde{a}^{\,n} \quad , \quad n \geq 0 \;\; avec \;\; \tilde{a} = \lambda.exp(i\beta) \qquad \text{(II .3)}$$

Où λ traduit l'atténuation, β la fréquence d'oscillation, γ l'ordre du filtre de n le temps échantillonné, ã est un coefficient complexe. Pour γ=4, on obtient la transformée en z suivante:

$$G(z) = \frac{\tilde{a}z^{-1}+4(\tilde{a}z^{-1})^2+(\tilde{a}z^{-1})^3}{(1-\tilde{a}z^{-1})^4} \qquad \text{(II .4)}$$

On simplifie alors cette fonction de transfert pour ne garder finalement que le dénominateur :

$$K(z) = \frac{1}{(1-\tilde{a}z^{-1})^4} \qquad \text{(II.5)}$$

Les pôles d'une fonction de transfert ayant généralement une influence bien plus importante sur la réponse en fréquence que les zéros pour les filtres à réponse impulsionnelle infinie (*Infinite Impulse Response*, IIR), toute laisse à penser que la réponse en fréquence ne sera que légèrement modifiée. On peut effectivement le vérifier sur la figure II.8 : ne pas

prendre en compte le numérateur de la fonction de transfert ne modifie que très peu la réponse en fréquence, et la plupart de ces différences sont situées dans la bande coupée. Par contre, comme les effets sur la réponse temporelle peuvent être plus prononcés, nous allons calculer l'erreur temporelle commise par cette approximation.

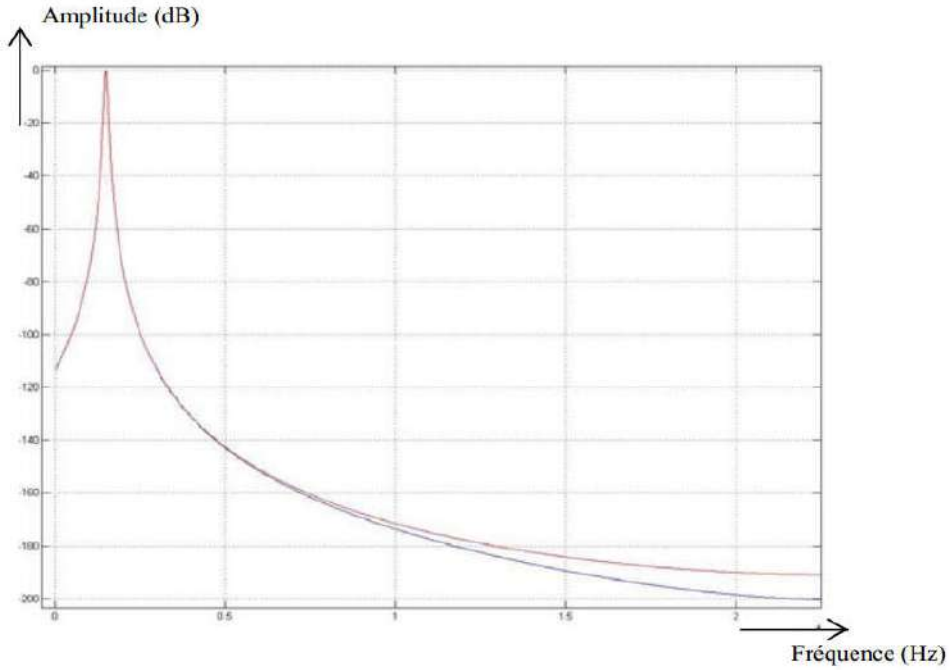


Figure II. 8. *Réponse en fréquence des fonctions G(z) (en bleu) et K(z) (en rouge) pour ã=0.9693 + 0.2104i, correspondant à une fréquence centrale de 1,5kHz et une largeur de bande de 50Hz.*

L'expression temporelle $k(n)$ de la fonction de transfert $K(z)$ est la suivante :

$$k(n) = \frac{\tilde{a}}{6}(n^3 + 6n^2 + 11n + 6) \qquad \text{(II .6)}$$

En négligeant le facteur 1/6, l'erreur absolue $e(n)$ et relative $r(n)$ de $k(n)$ par rapport à $g(n)$ est:

$$e(n) = 6.k(n) - g(n) = g(n)\left(\frac{6}{n} + \frac{11}{n^2} + \frac{6}{n^3}\right) \quad , n > 0$$

$$r(n) = \frac{e(n)}{g(n)} = \frac{6}{n} + \frac{11}{n^2} + \frac{6}{n^3} \qquad \text{(II .7)}$$

L'erreur relative étant en $1/n$, la forme et la structure temporelle fine du filtre Gammatone, qui ont une importance majeure pour la modélisation, notamment la caractéristique de phase de la cochlée, sont préservées par cette approximation tout-pôle[6].

II .6.2. Le filtre Gammatone dans le domaine temporel :

Le filtre Gammatone est défini dans le domaine temporel (la réponse impulsionnelle) par :

$$Gt(t) \propto t^{n-1} \exp(-2\pi bt) \cos(2\pi f_0 t + \emptyset), \qquad (t > 0) \qquad \text{(II .8)}$$

Il s'agit donc d'un filtre causal avec une réponse temporelle infinie. La forme de cette réponse impulsionnelle est d'une porteuse (*Tone*) de fréquence $f_o[Hz]$ modulée en amplitude, avec une enveloppe proportionnelle à $t^{n-1}\exp(-2\pi bt)$, qui est la distribution *Gamma*. Ces deux fonctions donnent naissance au nom *Gammatone* (de Boer et de Jongh, 1978).

Les paramètres du filtre *Gammatone* sont : l'ordre du filtre *n* ,qui (pour une valeur de *b* fixe*)* contrôle la forme relative de l'enveloppe qui devient oblique lorsque *n* augmente; *b* (en Hz) qui (pour *n* fixe*)* contrôle la durée de la réponse impulsionnelle, l'augmentation de *b* implique la diminution de la durée; f_0 (en Hz) qui détermine la fréquence de la porteuse; et $\emptyset$ (en radians), la phase du porteuse, Qui détermine la position relative de la structure fine du porteuse sur l'enveloppe[6].

II .6.3. Le filtre Gammatone dans le domaine fréquentiel :

Si *GT(f)* représente le filtre Gammatone dans le domaine fréquentiel (fonction de réponse en fréquence), alors :

$$GT(f) \propto \left[1 + \frac{j(f-f_0)}{b}\right]^{-n} + \left[1 + \frac{j(f+f_0)}{b}\right]^{-n}, \quad (-\infty < f < \infty) \qquad \text{(II .9)}$$

Cela peut être obtenu soit directement à partir de la définition dans le domaine temporel décrit ci-dessus, par l'application de la transformée de Fourier, soit en utilisant la propriété que le produit dans le domaine temporel d'une fonction Gamma et une fonction cosinus correspondra à la convolution dans le domaine fréquentiel de la transformée de Fourier $(1 + jf / b)^{-n}$ de la fonction Gamma avec une distribution à deux points à $\pm f_0$. Ici pour simplifier nous avons mis $\emptyset = 0$, car il n'a pas d'effet important sur les caractéristiques du filtre dans le domaine fréquentiel.

Le rôle des paramètres dans le domaine fréquentiel est clair dans la formule ci-dessus. f_0 est la fréquence centrale du filtre; Pour un ordre du filtre *n* fixe, *b* agit comme un paramètre d'échelle tel que le la bande passante du filtre augmente avec *b;* Le paramètre *n* contrôle la forme globale du filtre. Pour une valeur *b* fixe, la largeur de bande

diminue lorsque *n* augmente. *GT(f)* est approximativement symétrique par rapport à f_0 sur une échelle de fréquence linéaire.

Comme indiqué par de Boer et Krudenier (1988), le deuxième terme de l'équation (II .9) peut être ignoré lorsque f_0 / b est suffisamment grand - ce qui est toujours le cas lors de la modélisation du filtre auditif humain.

- donnant une fonction de réponse de fréquence approximative.

$$GT(f) \approx \left[1 + \frac{j(f-f0)}{b}\right]^{-n}, \qquad (0 < f < \infty) \qquad \text{(II .10)}$$

Pour les filtres Gammatone nécessaires à la modélisation du système auditif humain ($4 < f_0/b < 8$), Le deuxième terme de l'équation (II .9) est typiquement inférieur de 25 dB au niveau du premier terme, même à deux Bande passante en dessous de la fréquence centrale.

En raison de la *n-ième* forme de puissance de l'équation (II .10), le filtre Gammatone d'ordre *n* peut-être approximé par une cascade de *n* filtres Gammatone identiques d'ordre 1 avec la réponse fréquentielle :

$$\left[1 + \frac{j(f-f0)}{b}\right]^{-1} \qquad \text{(II .11)}$$

Chacun ayant la forme d'un filtre passe-bas à décalage de fréquence - un filtre de premier ordre qui a une application numérique récursive bien connue [7].

II .6.4.Largeur de bande rectangulaire équivalente du filtre Gammatone :

Lorsque fo / b est grand, la largeur de bande rectangulaire équivalente du filtre Gammatone est proportionnelle à *b*, et la constante de proportionnalité a_n ne dépend que de *n*.

Nous pouvons calculer la bande passante rectangulaire équivalente du filtre Gammatone en appliquant le théorème de Parseval. Si l'on suppose que $|GT(f)|^2$ est symétrique dans *f,* et que sa valeur maximale se produit lorsque $f = fo$ alors :

$$\int_{-\infty}^{+\infty} gt(t)^2 dt = \int_{-\infty}^{+\infty} |GT(f)|^2 df = 2|GT(f_0)|^2 ERB[GT] \qquad \text{(II .12)}$$

Où *ERB [GT]* est la bande passante rectangulaire équivalente du filtre Gammatone.

L'application de l'équation (II .12) au filtre Gammatone conduit à une intégrale traitable, ce qui donne la bonne approximation pour le constant a_n *:*

$$a_n = \frac{\pi(2n-2)!\, 2^{-(2n-2)}}{(n-1)!^2} \qquad \text{(II .13)}$$

La bande passante à 3dB d'un filtre Gammatone d'ordre *n* est $c_n b$ [Hz] où

$$c_n = 2\sqrt{2^{1/n} - 1} \qquad \text{(II .14)}$$

Le tableau (II .1) donne les valeurs de a_n et c_n pour le filtre Gammatone décrit ci-dessus.

***Tableau* II .1:** *les données pour le filtre Gammatone.*

n	a_n	a_n^{-1}	c_n	c_n^{-1}
1	3,141	0,318	2,000	0,500
2	1,570	0,637	1,288	0,777
3	1,178	0,849	1,020	0,980
4	0,982	1,019	0,870	1,149
5	0,889	1,164	0,772	1,296
6	0,773	1,293	0,700	1,429
7	0,709	1,411	0,646	1,550
8	0,658	1,520	0,602	1,662
9	0,617	1,621	0,566	1,767

Moore et Glasberg (1983) donnent la formule suivante pour la largeur de bande rectangulaire équivalente(en Hz) d'un filtre auditif humain d'une fréquence centrale f_0 [Hz] :

$$ERB[f_0] = 6{,}23 \times 10^{-6} {f_0}^2 + 93{,}39 \times 10^{-3} f_0 + 28{,}52 \quad (100 < f_0 < 10000) \qquad \text{(II.15)}$$

Ainsi, pour faire correspondre un filtre Gammatone d'ordre 4 et d'une fréquence centrale 1000 Hz, la valeur de b est ${a_n}^{-1} ERB\,[1000] = 1{,}019 \times 128{,}14 = 130{,}57$. La largeur de bande à 3dB de ce filtre est $c_n b = 0{,}870 \times 130{,}57 = 113{,}59\ Hz$[7].

II .6.5.Compensation de phase :

Lors de la représentation graphique de la sortie d'un banc de filtres, il est souvent utile d'aligner les pics des enveloppes des fonctions de la réponse impulsionnelle et d'aligner le pic de la structure fine avec le pic de chaque enveloppe.

Tout d'abord, les sommets des enveloppes peuvent être alignés en introduisant un facteur $t_c = (n - 1) / 2\pi b$ à la sortie du filtre. Deuxièmement, un pic dans la structure fine peut être déplacé vers le pic de l'enveloppe en introduisant une correction de phase $\emptyset_c = -2\pi f_0 t_c$.Cela conduit au filtre (non causal)

$$\tilde{g}t(t) \propto (t + t_c)^{n-1} exp[-2\pi b(t + t_c)] cos(2\pi f_0 t)(t \geq -t_c) \qquad \text{(II .16)}$$

Qui aligne le pic de la réponse impulsionnelle de tous les filtres à t =0 [7].

II .7.Conclusion:

Dans ce chapitre, nous avons vu Le filtre de Gammatone présenté par Johannesma pour décrire la réponse de la cochlée en 1972, qui a été adopté comme une base dans certain nombre de travaux réussis de modélisation du système auditifs humain.

III.1. INTRODUCTION:

Dans ce chapitre, nous nous présentons les résultats de simulation d'un algorithme de séparation de sources. Dans notre projet, nous avons utilisé le logiciel MATLAB.

Dans une première partie, nous commençons à expliquer le principe du banc de filtre Gammatone en générant sa réponse implusionelle et en calculant son cochléagramme. Une deuxième partie sert utiliser le banc de filtres, dans l'algorithme MGD IS-NMF2D afin d'estimer les sources à partir d'un mélange.

Les résultats obtenus pour la séparation de sources sont présentés à la fin de ce chapitre.

III.2. Qu'est-ce que MATLAB ?

MATLAB pour MATtrix LABoratory, est une application qui a été e conçue afin de fournir un environnement de calcul matriciel simple, efficace, interactif et portable, permettant la mise en œuvre des algorithmes développés dans le cadre des projets *linpack*et*eispack*.

MATLAB est constitué d'un noyau relativement réduit, capable d'interpréter puis d'évaluer les expressions numériques matricielles qui lui sont adressées :

– soit directement au clavier depuis une fenêtre de commande ;
– soit sous forme de séquences d'expressions ou **scripts** enregistrées dans des fichiers-texte appelés *m-files* et exécutées depuis la fenêtre de commande ;
– soit plus rarement sous forme de fichiers binaires appelés *mex-files* ou fichiers .mex générés à partir d'un compilateur C ou fortran.

Ce noyau est complété par une bibliothèque de fonctions prédéfinies, très souvent sous forme de fichiers m-files, et regroupes en paquetages ou toolboxes. A cote des toolboxes requises local et Matlab, il est possible d'ajouter des toolboxes spécifiques à tel ou tel problème mathématique, Optimization Toolbox, Signal Processing Toolbox par exemple ou encore des toolboxes crées par l'utilisateur lui-même. Un système de chemin d'accès ou path permet de préciser la liste des répertoires dans lesquels MATLAB trouvera les différents fichiers m-files.

III.2.1. Le son sous Matlab :

Windows propose un standard des fichiers son: Le format '.wav' ('.au' est le standard chez unix). Ce format est rudimentaire et non compressé. Comme tout fichier formaté, un fichier '.wav' comporte une entête : Champs de 44 octets, suivie des données (le son proprement dit).

III.2.2. Lecture d'un fichier son :

Pour lire un fichier '.wav' dans Matlab, nous devons exécuter la commande suivante :

[x,fs]=wavread('pianotrump.wav') ;

Où ***x*** est un vecteur colonne; et ***fs*** est la fréquence d'échantillonnage en Hertz.

III.2.3. Ecriture d'un fichier son:

La commande suivante :

wavwrite(x,fs,'son.wav') ;

Permet de crée un fichier son a un nom « ***son.wav*** » dont les échantillons (compris entre -1 et 1) se trouvent dans le vecteur *x* avec une fréquence d'échantillonnage ***fs***.

III.3. Les données de traitement :

Dans notre projet, un fichier audio contenant le mélange de sons de deux instruments de musique (piano et trumpet), a été lu sous la forme suivante :

[x,fs]=wavread('pianotrump.wav') ;

x représente le signal d'entrée qui contient le mélange de deux signaux et ***fs*** est la fréquence d'échantillonnage en Hertz.

Pour réaliser notre banc de filtres Gammatone, nous prenons :

- Nombre de canaux=128 implique N=2^7 qui représente la taille de signal.
- Une bande fréquentielle de travail entre 50 et 8000 Hz (le spectre audible d'une oreille saine est compris entre 20 et 20000Hz).

La figure suivante montre le tracé échantillonné du signal de mélange choisi.

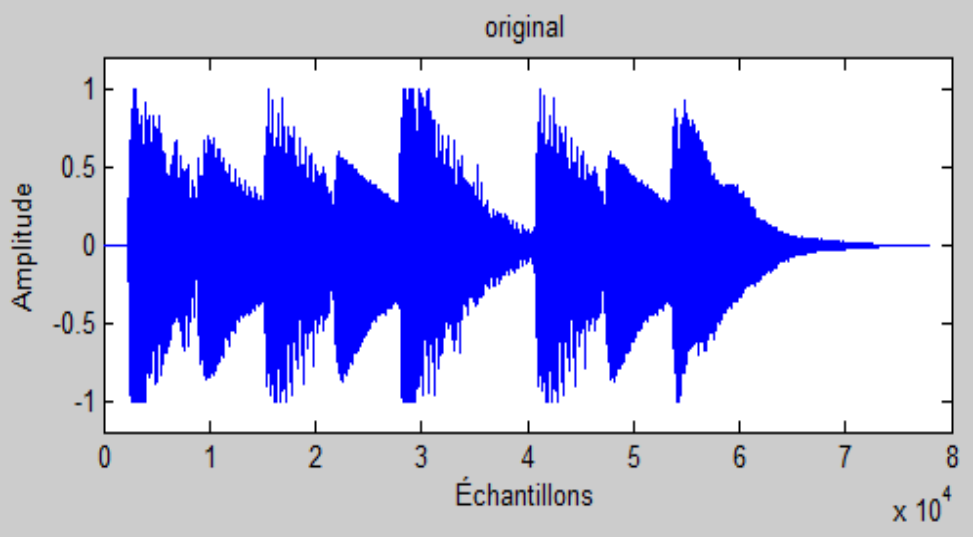


Figure III. 1. Le tracé du mélange de deux signaux à séparés.

III.4. Banc de filtres Gammatone :

La réponse impulsionnelle du filtre auditif est représentée par une fonction Gammatone définie dans le domaine temporel par l'expression suivante :

$$g(t) = At^{n-1}\exp(-2\pi bB(fc)t)\cos(-2\pi fct+\varphi) \qquad \text{(III.1)}$$

Où ***A*** est un paramètre de normalisation d'amplitude, ***n*** est l'ordre du filtre, ***fc*** est la fréquence centrale du filtre, ***B*** est la largeur des bandes passantes des filtres et ***bB(fc)*** un paramètre définissant l'enveloppe du filtre.

III.5. Choix de l'échelle fréquentielle :

Les bandes passantes du banc de filtres, supposés être idéaux, doivent être disjointes mais leur union doit couvrir la bande utile du signal. La largeur de ces bandes peut être uniforme, ou non uniforme, elle suit dans ce cas, les caractéristiques des bandes critiques, la formule qui permet de convertir une fréquence *f* en Hertz de sa valeur *ERB* est :

$$cf=(10.^{(erb/21.4)}-1)/4.37e-3 \qquad \text{(III.2)}$$

Le système auditif humain ne perçoit pas les signaux selon une échelle uniforme. Notamment, il effectue une analyse plus précise dans les fréquences graves que dans les bandes hautes du spectre. Ainsi, réalisant l'analyse spectrale selon une échelle en

fréquence conforme à notre perception auditive, l'échelle *ERB* est la largeur des bandes critiques et elle modélise la résolution fréquentielle des filtres auditifs. La relation moyenne pour plusieurs sujets "normaux"(à propos de l'audition) entre la fréquence centrale *f* en Hz et l'*ERB* en Hz est la suivante :

$$b = 1.019*24.7*(4.37*cf/1000+1) \qquad \text{(III.3)}$$

III.6. Résultats :

La figure III.2 nous donne la représentation fréquentielle d'un banc de filtres Gammatone 12 bandes, de fréquence minimale 70Hz.

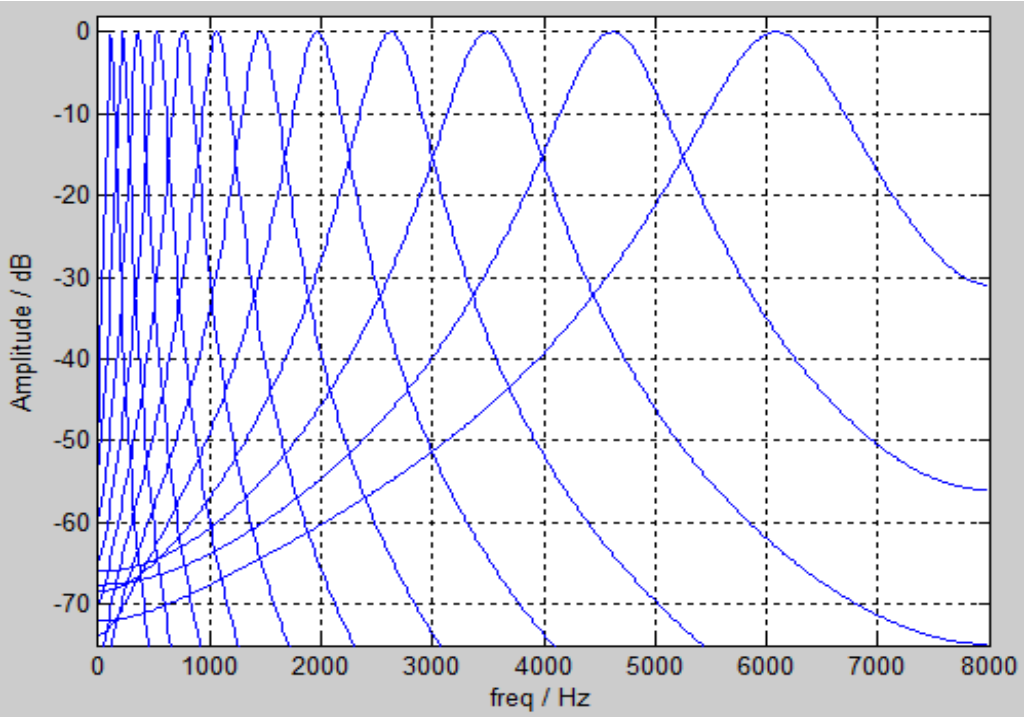


Figure III. 2. *Réponse en fréquence d'un banc de filtre Gammatone 12 bandes, f* min *=70Hz.*

Le code Matlab qui permet de générer la réponse implusionnelle de filtre Gammatone est le suivant :

```
gt = zeros(N_can,gL);  % initialisation
tmp = [1:gL]/Fs; % crée le vecteur de temps
for i = 1:N_can
gain = 10^((loudness(cf(i))-60)/20)/3*(2*pi*b(i)/Fs).^4; % loudness-based
    gain adjustments
gt(i,:) = gain*Fs^3*tmp.^(filOrdre-1).*exp(-2*pi*b(i)*tmp)…
.*cos(2*pi*cf(i)*tmp+phase(i));
end
```

La fonction "***loudness***" dans le code précédant permet d'ajuster le signal à l'effet de l'oreille moyenne selon le standard *ISOBS3383*.

La figure suivante, montre la réponse impulsionnelle du canal 3.

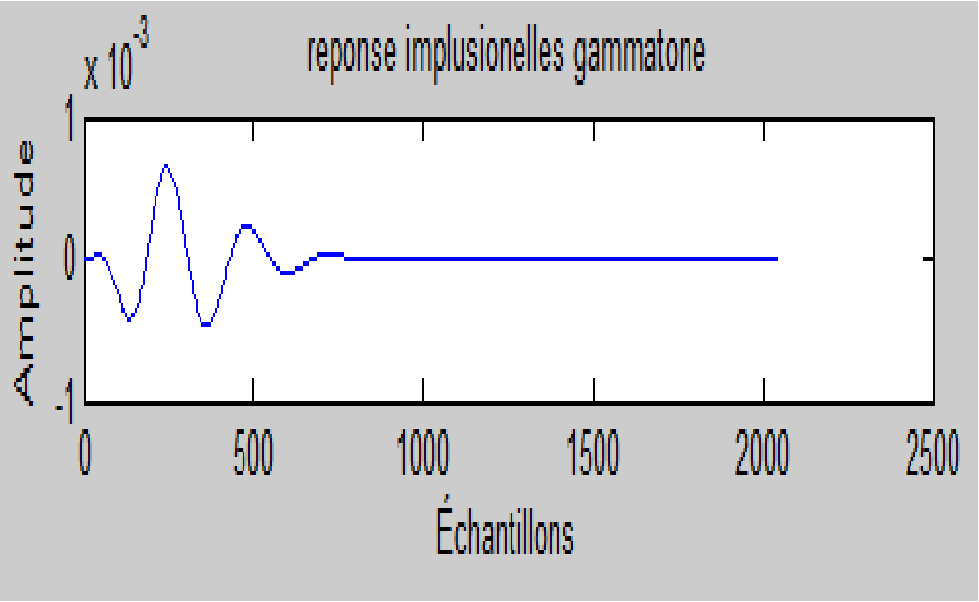


Figure III. 3. *Réponse implusionelle de la bande passante N°3. Les paramètres sont : b3 =32.3479et cf3 = 65.2661*

III.7. Transformée de Fourier discrète :

Après la génération des réponses impulsionnelles des filtres Gammatone, nous devons filtrer le signal du mélange par le banc de filtres en utilisant la commande *fftfilt* du Matlab.

```
% convertir signal a un vecteur colonne
sign = reshape(x,sigLong,1);
 % filtragegammatoneutilisant FFTFILT
gf = fftfilt(gt',repmat(sign,1,N_can))';
```

La figure suivante, montre le signal de sortie du 3ème filtre dans le banc réalisé, après filtrage :

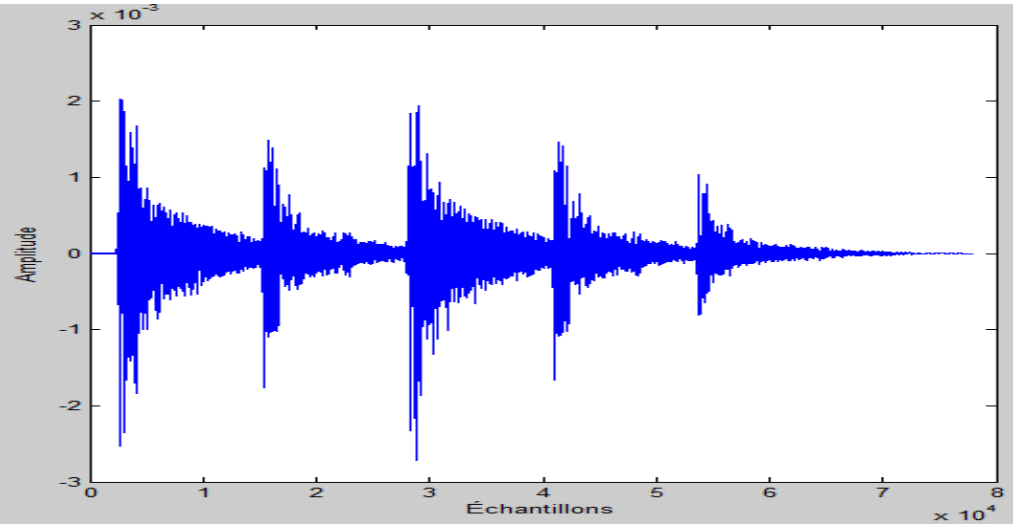


Figure III. 4. Le *Signal de filtre gammatone filtré de la bande passante N°3.*

III.8. Le cochléagramme :

Le code Matlab permettant de générer le cochléogramme à partir des réponses de banc de filtres Gammatone est décrit dans le code suivant :

```
winLength = 320;      % Longueur de fenêtre par défaut dans les  %points
    d'échantillonnage qui est de 20 ms pour une fréquence
    %d'échantillonnage de 16 KHz
[N_can,sigLong] = size(gf);  % Nombre de canaux et longueur du signal
winShift = winLength/2;% décalage dans les trames (par défaut est %demi
    trame)
increment = winLength/winShift;   % increment dans les trames
M = floor(sigLong/winShift);     % Nombre de périodes
% Calcule de l'énergie pour chaque trame dans chaque canal
a = zeros(N_can,M);
for m = 1:M
for i = 1:N_can
if m < increment
a(i,m) = gf(i,1:m*winShift)*gf(i,1:m*winShift)';
else
startpoint = (m-increment)*winShift;

    a(i,m)=gf(i,startpoint+1:startpoint+winLength)*gf(i,startpoint+1:star
    tpoint+winLength)';
end
end
end
```

La figure suivante montre la variation de l'énergie du signal à la sortie de la 3ème bande.

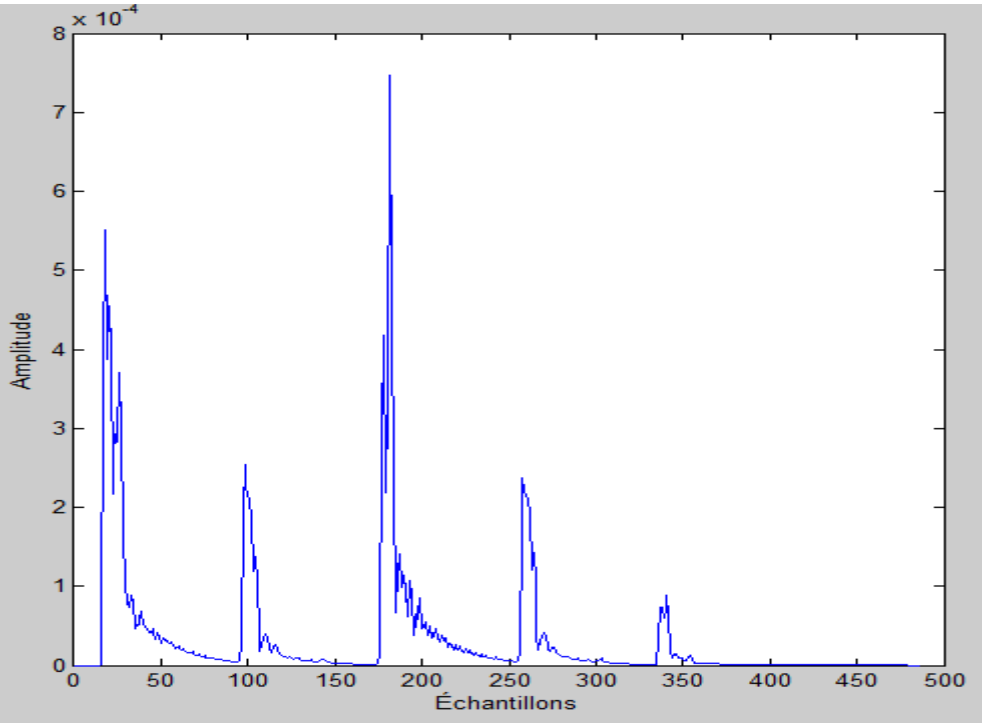


Figure III. 5. La sortie de l'énergie *d'un filtre Gammatone de la bande passante N°3. Les paramètres sont :b3 =32.3479et cf3 = 65.2661*

III.9. Algorithme de séparation de sources :

L'algorithme MGD IS-NMF2D est basé sur la descente de gradient multiplicatif. L'algorithme est décrit comme suit :

- **Entrée:** $|Y|^2$, matrices non négatives aléatoires D^{τ} etH^{φ},φ,τ.
- **Sortie:** D^{τ} et H^{φ}
- **Procédure:**

Calculer la valeur initiale de la fonction de coût 'cost' en utilisant la fonction suivant :

$$-\log p(Y|D,H) \doteq d_{IS}(|Y|^{.2}_{f,ts} | \sum_{i=1}^{I} \sum_{\tau=0}^{\tau max} \sum_{\varphi=0}^{\varphi max} D^{\tau}_{f-\varphi,i} H^{\varphi}_{i,ts-\tau}) \quad \text{(III.4)}$$

✓ Itérer pour n=1 : Nombre maximum d'itérations

- Calculer $Z = \sum_{\tau} \sum_{\varphi} D^{\tau}_{f-\varphi} H^{\varphi}_{ts-\tau}$

- Mise à jour $D^{\tau}_{f',i'}$ en utilisant (III.5) pour tous les τ, φ.

$$D^{\tau'}_{f',i'} \leftarrow D^{\tau'}_{f',i'} \frac{\sum_{\varphi,ts}(Z_{f'+\varphi,ts})^{-2} \left|Y\right|^{2}_{f'+\varphi,ts} H^{\varphi}_{i',ts-\tau'}}{\sum_{\varphi,ts}(Z_{f'+\varphi,ts})^{-1} H^{\varphi}_{i',ts-\tau}} \quad \text{(III.5)}$$

- Normaliser $D^{\tau}_{f',i'}$.
- Calculer $Z = \sum_{\tau}\sum_{\varphi} D^{\tau}_{f-\varphi} H^{\varphi}_{ts-\tau}$
- Mise à jour $H^{\varphi'}_{i'-ts'}$ en utilisant (III.6) pour tous les τ, φ.

$$H^{\varphi'}_{i',ts'} \leftarrow H^{\varphi'}_{i',ts'} \frac{\sum_{\varphi,ts}(Z_{f,ts'+\tau})^{-2} \left|Y\right|^{2}_{f,ts'+\tau} D^{\tau}_{f-\varphi',i'}}{\sum_{\tau,f} D^{\tau}_{f-\varphi',i'}(Z_{f,ts'+\tau})^{-1}} \quad \text{(III.6)}$$

- Normaliser $H^{\varphi'}_{i'-ts'}$.
- Calculer la valeur '*cost*' en utilisant la fonction (III.4).

- Fin : Critère d'arrêt $\frac{cost(n-1)-cost(n)}{cost(n)} < \varepsilon$

III.10. Estimation des sources :

Les deux matrices dont nous cherchons à séparer à partir de $\left|Y_{f,ts}\right|^{2}$ Sont $\left|\hat{X}1(f,ts)\right|^{.2}$ et $\left|\hat{X}2(f,ts)\right|^{.2}$.

Ces matrices sont estimées en utilisant les deux formules suivantes :

$$\left|\hat{X}1(f,ts)\right|^{.2} = \sum_{\tau=0}^{\tau max}\sum_{\varphi=0}^{\varphi max} D^{\tau}_{f-\varphi,1}\, H^{\varphi}_{1,ts-\tau} \quad \text{(III.7)}$$

$$\left|\hat{X}2(f,ts)\right|^{.2} = \sum_{\tau=0}^{\tau max}\sum_{\varphi=0}^{\varphi max} D^{\tau}_{f-\varphi,2}\, H^{\varphi}_{2,ts-\tau} \quad \text{(III.8)}$$

Qui sont ensuite utilisés pour générer le masque binaire :

$$\begin{cases} mask_i(f,ts) = 1 \; si \; \left|\hat{X}i(f,ts)\right|^{.2} > \left|\hat{X}j(f,ts)\right|^{.2} \\ mask_i(f,ts) = 0 \quad si \; non \end{cases} \quad \text{(III.9)}$$

Enfin, les sources estimées du domaine temporel sont obtenues Comme $\hat{X}i = Resynthesize(mask_i \bullet Y)$ pour i = 1,2 où $\hat{x}i = [\hat{x}i(1), \ldots, \hat{x}i(T)]^{T}$ Désigne la ième source estimée.

Les sources estimées du domaine temporel sont resynthétisées en pondérant le mélange de cochléagramme Par le masque et la correction des changements de phase introduits pendant le filtrage Gammatone.

Le code Matlab permettant de synthétiser à partir d'un signal de mélange et d'un masque binaire les deux sources séparément est donné par :

```
for c = 1:N_can
    temp1(c,:) = fliplr(temp1(c,:))/midEarCoeff(c);    % Temps de %sortie
    du filtre inversé et normaliser les coefficients de l'oreille
    %moyenne
    temp2 = fftfilt(gt(c,:),temp1(c,:));
temp1(c,:) = fliplr(temp2(1:sigLong))/midEarCoeff(c);    % Le %temps
    revient à nouveau et normalisation des coefficients de %l'oreille
    moyenne
weight = zeros(1, sigLong);       % Calculer la pondération
for m = 1:numFrame-increment/2+1      % La valeur du masque peut      %
    être binaire ou rationnelle
startpoint = (m-1)*winShift;
if m <= increment/2
weight(1:startpoint+winLength/2) = . . . weight(1:startpoint+winLength/2)
    + mask(c,m)*coswin(winLength/2-startpoint+1:end);
else
weight(startpoint-winLength/2+1:startpoint+winLength/2) = . . .
    weight(startpoint-winLength/2+1:startpoint+winLength/2) + . . .
    mask(c,m)*coswin;
end
end
```

III.11. Résultats de simulation :

Le banc de filtre Gammatone, associé à l'algorithme de synthèse permet d'effectuer

des traitements audio. Le son est analysé et décomposé par bande de fréquence, modifié, puis resynthétisé. Les figures suivantes nous donnent les deux sources séparées du mélange.

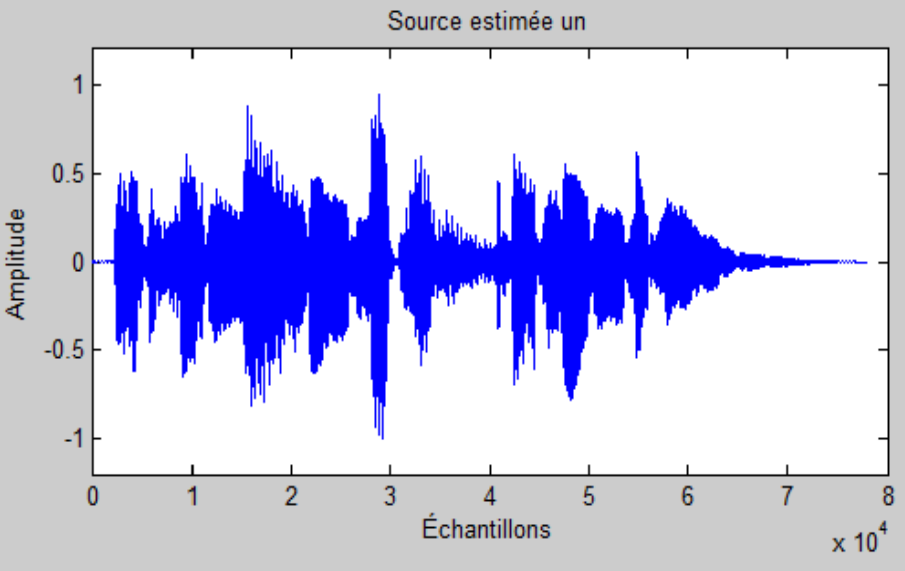


Figure III. 6. La première Source estimée.

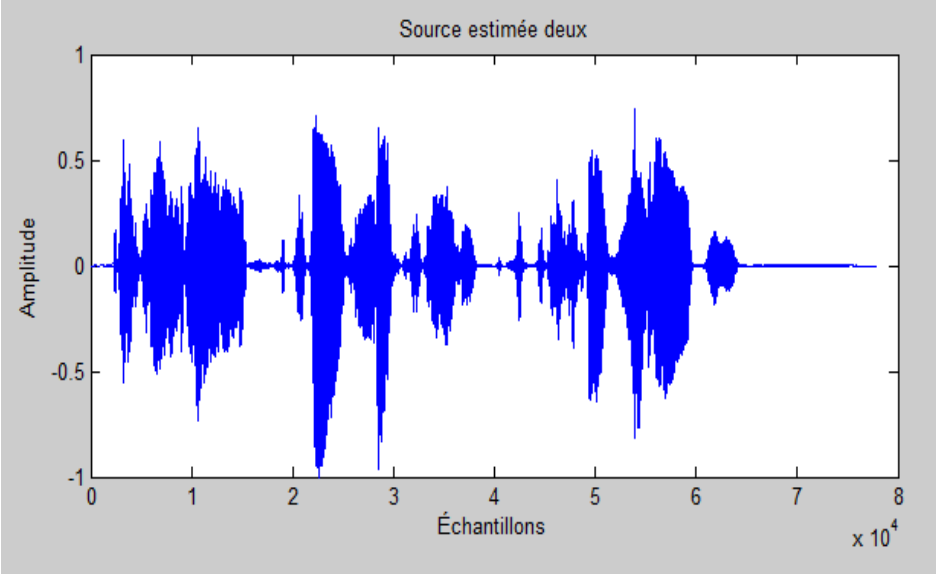


Figure III. 7. La deuxième Source estimée.

III.12. Conclusion :

Dans ce chapitre, nous avons présenté la méthode de séparation Aveugle de source audio, en utilisant l'algorithme MGD IS-NMF2D qui est basée sur le banc de filtres Gammatone.

Cette méthode montre son efficacité en point de vu résultats obtenus, et si entend les signaux séparés on peut les connaitre facilement.

CONCLUSION GENERALE :

Dans ce mémoire, Nous avons étudié et simulé une nouvelle méthode pour séparer deux sources audio provenant d'un canal unique. En outre, l'algorithme de factorisation bidimensionnelle de matrices non négatives optimisée a été présenté: MGD IS-NMF2D.

Associé au support théorique de la séparabilité des signaux dans le domaine TF, le système de séparation utilisant le banc de filtres Gammatone a montré un Succès considérable. La méthode proposée bénéficie d'au moins trois Avantages importants:

- Tout d'abord, il évite de fortes contraintes de séparation de sources sans connaissances d'information apriori.
- Deuxièmement, le cochléagramme rendu par le banc de filtres Gammatone a une résolution TF non uniformes, et qui permet de séparer le signal mixte et d'améliorer ainsi l'efficacité de la séparation des sources.
- Finalement, La méthode détient une propriété souhaitable d'invariance à l'échelle qui permet aux composants à faible énergie dans le cochléagramme d'avoir la même importance relative que les énergies élevées.

Comme perspectives, nous proposons d'utiliser le principe de cet algorithme dans la séparation de plus de deux sources, vue les résultats que nous avons obtenus et qui montrent l'efficacité de cette méthode.

Bibliographie

[1] : Yasmina BENABDERRAHMANE, `` SEPARATION AVEUGLE DE SIGNAUX DE PAROLE UTILISANT LES STATISTIQUES D`ORDRE SUPRIEUR ET LA DECONPOSISION EN SOUS-ESPACES`` , Benabderrahmane_2011.

[2] : Sebastien Lagrange;``Contributions aux méthodes d'estimation en Aveugle``, Thèse Lagrange Sébastien le 01 decembre 2005.

[3] : Simon ARBERET ; ``Estimation robuste et apprentissage aveugle de modèles pour la séparation de sources sonores``, arberet2008.

[4] : Abdeldjalil AISSA-EL-BEY ;``séparation aveugle de source audio``seminaire_aissa_el_bey_08_11_07.

[5] : Hakim BOUMARAF ; ``SEARATION AVEUGLE DE MELANGES CONVOLUTIFS DE SOURCE ``, Boumaraf Thesis 26 octobre 2005.

[6] : Dan Gnansia ;``Modèle auditif en temps réel ``, Gnansia juillet 2005.

[7] : John Holdsworth ;``Implementing a GammaTone Filter Bank*``, SVOSAnnex C 1988.

[8] : L'Université Pierre et Marie Curie ; ``Audition active et intégration sensorimotrice pour un robot autonome bioinspiré ``, thèse _archivage_2969147 15 mai 2014.