

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITÉ ABDERRAHMANE MIRA – BEJAIA



FACULTÉ DES SCIENCES EXACTES
DÉPARTEMENT DE MATHÉMATIQUES

Mémoire de Fin d'Études

Présenté pour l'obtention du diplôme de

Master en Mathématiques

Option : Mathématiques de l'Intelligence Artificielle" M.I.A "

Thème du mémoire

*Réduction du bruit par la transformée de Fourier
et Machine Learning dans les aides auditives*

Présenté par :

Mlle. KASSA Celina

Soutenu le 29 juin 2025 devant le jury composé de :

| | | | |
|----------------------------|-------|---------------|-------------------|
| Mr. BOUZIDI L'hadi | MCB | Président | U. A. Mira Béjaia |
| Mme. BOULAHIA-TALBI Fatiha | Prof. | Rapporteur | U. A. Mira Béjaia |
| Mme. REBOUH Nadjette | MAA | Co-rapporteur | U. A. Mira Béjaia |
| Mme. MEZZAH Samia | MAA | Examinatrice | U. A. Mira Béjaia |
| Mr.KHOUFACHE Samir | MAA | Examineur | U. A. Mira Béjaia |

Année universitaire : 2024 – 2025

Remerciements

Je souhaite tout d'abord exprimer ma profonde gratitude à Dieu, dont la guidance et la sagesse m'ont accompagnée tout au long de ce parcours. Sa présence constante a été une source de force et d'inspiration dans les moments de doute.

Je tiens également à adresser mes plus sincères remerciements à mes encadrantes, Madame **BOULAHIA TALBI Fatiha** et Madame **REBOUH Nadjette**. Leur expertise, leur accompagnement bienveillant et leurs conseils avisés ont été des soutiens indispensables tout au long de ce projet. Leur dévouement et leur patience ont nourri ma motivation et m'ont guidée avec rigueur et bienveillance à chaque étape de mon travail.

Je tiens à exprimer mes sincères remerciements à Monsieur **BOUZIDI L'hadi**, président du jury, ainsi qu'à Monsieur **KHOUFACHE Samir** et Madame **MEZZAH Samia**, membres du jury. Je leur suis profondément reconnaissante d'avoir accepté d'évaluer ce travail et d'avoir honoré cette soutenance de leur présence.

Je remercie profondément mes parents, dont l'amour inconditionnel, les sacrifices et la confiance ont toujours été le socle de ma persévérance. Leur soutien représente la base solide de mon parcours académique et personnel.

Je tiens à exprimer ma profonde gratitude à mon très cher ami **SALHI Malek**, dont le soutien, les conseils et la bienveillance ont été d'une grande valeur tout au long de ce travail. Son écoute attentive, sa disponibilité et ses encouragements constants ont grandement contribué à la réussite de ce projet. Merci pour ta présence précieuse à chaque étape.

Je souhaite également adresser un remerciement sincère à l'ensemble de mes professeurs rencontrés tout au long de mon parcours universitaire. Chacun, à sa manière, a contribué à construire mes connaissances, à façonner ma pensée et à nourrir mon intérêt pour les sciences.

Enfin, un remerciement chaleureux à toute la promotion de Mathématiques de l'Intelligence Artificielle, pour les moments de partage, d'entraide et de camaraderie vécus ensemble. Ce chemin a été d'autant plus riche grâce à vous. À ma famille et à mes amis, je suis infiniment reconnaissante pour votre compréhension, vos encouragements et votre présence à mes côtés.

Dédicaces

Je dédie ce Projet de Fin d'Études à mes parents, véritables piliers de mon parcours. Leur amour inconditionnel, leurs sacrifices constants et leur foi inébranlable en mes capacités ont été la lumière qui a dissipé mes doutes et la force qui m'a incitée à persévérer, même dans les moments les plus difficiles. Ce succès leur revient autant qu'à moi.

Je tiens à exprimer ma sincère gratitude à mes encadrantes, dont l'accompagnement attentif, la bienveillance et les conseils avisés ont joué un rôle essentiel dans la réalisation de ce travail. Leur rigueur scientifique et leur engagement constant ont été une source d'inspiration et de motivation tout au long de ce projet.

Je remercie également tous mes professeurs pour leur expertise et leur dévouement. Leur enseignement a non seulement élargi mes connaissances, mais a également éveillé ma curiosité intellectuelle et affiné mes compétences.

Je dédie également ce travail à ma promotion de Mathématiques de l'Intelligence Artificielle, avec qui j'ai partagé des moments d'entraide, de réflexion et de complicité. Ce parcours a été riche d'expériences humaines et académiques grâce à vous.

Enfin, à ma famille et à mes amis : votre soutien constant, vos encouragements sincères et votre patience face aux défis de ce parcours ont été pour moi une source inestimable de réconfort. Votre présence à mes côtés a donné tout son sens à cette aventure.

Table des matières

| | |
|---|-----------|
| Liste des figures | 6 |
| Liste des tableaux | 7 |
| Introduction Générale | 8 |
| 1 L'oreille humaine : Structure, pathologies et solutions technologiques | 10 |
| 1.1 Intorduction | 10 |
| 1.2 Anatomie de l'oreille humaine | 10 |
| 1.3 Les maladies de l'oreille | 13 |
| 1.3.1 La surdit  | 14 |
| 1.4 Proth ses auditives | 15 |
| 1.4.1 Les composants principaux d'un appareil auditif | 16 |
| 1.4.2 Fonctionnement technique d'une aide auditive | 17 |
| 1.5 Conclusion | 17 |
| 2 Fondements du traitement de signal | 19 |
| 2.1 Introduction | 19 |
| 2.2 D finitions et exemples | 19 |
| 2.2.1 Le signal sonore : Nature et propri t s physiques | 20 |
| 2.3 Repr sentation et classification des signaux | 22 |
| 2.3.1 Repr sentation des signaux | 22 |
| 2.3.2 Classification des signaux | 22 |
| 2.4 Nature, classification et traitement du bruit | 29 |
| 2.4.1 Impact du bruit sur la perception | 30 |
| 2.4.2 Calssification et caract ristiques du bruit | 30 |
| 2.4.3 Mod lisation du bruit dans un signal audio | 31 |
| 2.5 Bruits courants et principes de r duction associ s | 31 |
| 2.5.1 M thodes de r duction du bruit : revue des techniques | 32 |
| 2.5.2 Indicateurs de qualit  d'un signal audio | 32 |
| 2.6 Conclusion | 33 |

| | | |
|----------|--|-----------|
| 3 | La Transformée de Fourier dans le traitement du signal | 35 |
| 3.1 | Introduction | 35 |
| 3.2 | Pourquoi le domaine fréquentiel? | 35 |
| 3.3 | Transformée de Fourier "TF" d'un signal continu | 36 |
| 3.3.1 | Définition et propriétés | 36 |
| 3.4 | Transformée de Fourier Discrète | 40 |
| 3.4.1 | Définitions et exemples | 40 |
| 3.4.2 | La transformée de Fourier discrète inverse | 41 |
| 3.4.3 | Propriétés de la TFD | 42 |
| 3.4.4 | Interprétation matricielle de la TFD et de la TFDi | 43 |
| 3.4.5 | Transformée de Fourier rapide | 45 |
| 3.5 | Transformée de Fourier à court terme : Approche fréquentielle et segmentation temporelle | 46 |
| 3.5.1 | Principes de la transformée de Fourier à court terme | 46 |
| 3.5.2 | Fenêtrage et analyse spectrale | 47 |
| 3.5.2.1 | Fenêtre rectangulaire | 47 |
| 3.5.2.2 | Fenêtres de Hamming et Hanning | 47 |
| 3.5.3 | Fenêtrage temporel et définition de la TFCT | 48 |
| 3.5.4 | Définition mathématique de la TFCT discrète | 48 |
| 3.5.5 | Écriture matricielle de la TFCT | 49 |
| 3.6 | Conclusion | 51 |
| 4 | Réduction de Bruit par la soustraction spectrale : Classique et intelligente | 52 |
| 4.1 | Introduction | 52 |
| 4.1.1 | Principe de la soustraction spectrale | 52 |
| 4.1.2 | Variantes de la soustraction spectrale | 54 |
| 4.1.2.1 | Soustraction spectrale avec sur-soustraction (α -SS) | 54 |
| 4.1.2.2 | Soustraction spectrale multi-bandes | 55 |
| 4.1.2.3 | Soustraction spectrale probabiliste (MMSE-SS) | 55 |
| 4.1.3 | Applications pratiques de la soustraction spectrale | 55 |
| 4.1.4 | Intégration des algorithmes de réduction de bruit dans les aides auditives | 56 |
| 4.1.4.1 | Défis spécifiques | 57 |
| 4.1.4.2 | Modèles d'estimation du bruit | 57 |
| 4.1.4.3 | Approches modernes hybrides | 57 |
| 4.1.4.4 | Exemple d'implémentation | 57 |
| 4.2 | Implémentation et évaluation des performances | 57 |
| 4.2.1 | Structure générale de l'implémentation | 58 |
| 4.2.2 | Métriques de performance | 59 |
| 4.2.3 | Approche 1 : Soustraction spectrale à estimation statique | 61 |
| 4.2.4 | Approche 2 : Soustraction spectrale avec VAD énergétique | 63 |
| 4.2.5 | Approche 3 : Soustraction spectrale adaptative avec VAD basé sur SVM | 65 |
| 4.2.6 | Comparaison des approches de réduction de bruit | 68 |

| | |
|----------------------------|-----------|
| 4.3 Conclusion | 69 |
| Conclusion Générale | 70 |
| Bibliographie | 72 |

Listes des abréviations et des symboles

Liste des Abréviations

CAN : Convertisseur Analogique-Numérique

dB : Décibel

dB SPL : Decibel Sound Pressure Level

f : Fréquence d'un signal

f_e : Fréquence d'échantillonnage

f_{\max} : Fréquence maximale d'un signal

Hz : Hertz

LTI : Linear Time-Invariant ou (linéaire et invariant dans le temps)

MSE : Erreur quadratique moyenne (Mean Squared Error)

OMS : Organisation Mondiale de la Santé

PESQ : Évaluation perceptuelle de la qualité de la parole (Perceptual Evaluation of Speech Quality)

RSB : Rapport Signal sur Bruit (Signal-to-noise ratio, SNR)

SDR : Rapport signal-distorsion (Signal-to-Distortion Ratio)

SNR : Signal-to-Noise Ratio

SS : Soustraction spectrale

SVM : Support Vector Machine

T_e : Période d'échantillonnage

TFR : Transformée de Fourier Rapide (FFT : Fast Fourier Transform)

TF : Transformée de Fourier

TFCT : Transformée de Fourier à Court Terme

TFCTI : Transformée de Fourier à Court Terme Inverse

TFD : Transformée de Fourier Discrète (DFT : Discrete Fourier Transform)

TFDI : Transformée de Fourier Discrète Inverse

VAD : Détection de la parole (Voice Activity Detection)

Liste des Symboles

α : Facteur de sursoustraction

β : Facteur de limitation des artefacts

$b(t)$: Bruit additif

$B[k]$: Spectre complexe du bruit

$|B[k]|$: Magnitude du spectre du bruit

c_n : Coefficients exponentiels de la série de Fourier

δ : Distribution de Dirac

i : Unité imaginaire

$p(x)$: Densité de probabilité d'un signal

$s(t)$: Signal utile

$S[k]$: Spectre complexe du signal utile

$|S[k]|$: Magnitude du spectre du signal utile

$x(t)$: Signal continu

$x_e(t)$: Signal échantillonné

$x[n]$: Signal temporel discret

$X[k]$: Spectre complexe du signal bruité

$|X[k]|$: Magnitude du spectre du signal bruité

$X[m, k]$: DFT de la m-ième trame

$|\hat{B}[k]|$: Estimation de la magnitude du bruit

$|\hat{S}[k]|$: Estimation de la magnitude du signal utile

$w[n]$: Fonction de fenêtre

Table des figures

| | | |
|------|--|----|
| 1.1 | Anatomie de l'oreille | 11 |
| 1.2 | Cellules ciliées | 12 |
| 1.3 | La cochlée | 12 |
| 1.4 | La surdit  | 15 |
| 1.5 | Sch ma de l'appareil auditif | 15 |
| 1.6 | Microphones omnidirectionnels | 16 |
| 1.7 | Microphones directionnels (bidirectionnels) | 16 |
| 1.8 | Les composants principaux d'un appareil auditif | 17 |
| 2.1 | Variation de la pression de l'air dans le temps | 20 |
| 2.2 | Repr sentation de l'amplitude d'un signal sonore | 20 |
| 2.3 | La fr quence | 21 |
| 2.4 | Exemples de spectres du son | 21 |
| 2.5 | Signal analogique | 23 |
| 2.6 | Signal triangle | 23 |
| 2.7 | Impulsion de Dirac | 24 |
| 2.8 | Signal num rique | 24 |
| 2.9 | Sch ma d' chantillonnage d'un signal | 26 |
| 2.10 | Echantillonnage d'un signal analogique | 26 |
| 2.11 | Restitution d'un signal num ris  | 27 |
| 2.12 | Signaux d terministes | 29 |
| 3.1 | Joseph Fourier | 35 |
| 3.2 | Le signal $x(t) = e^{-a t }$ et sa transform e de Fourier | 37 |
| 3.3 | Pr sentation d'un syst me LTI | 39 |
| 3.4 | TFD et TFDi | 41 |
| 4.1 | Diagramme de la soustraction spectrale | 54 |
| 4.2 | Applications pratiques de la soustraction spectrale | 56 |
| 4.3 | Cha ne de traitement audio commune aux deux algorithmes. | 59 |
| 4.4 | Aper u concret de l'utilisation s quentielle des fonctions | 60 |
| 4.5 | Approche 1 : Spectrogramme du signal bruit  | 62 |
| 4.6 | Approche 1 : Spectrogramme du signal du bruit estim  avec VAD statique | 62 |

| | | |
|------|--|----|
| 4.7 | Approche 1 : Spectrogramme du signal nettoyé avec VAD statique | 62 |
| 4.8 | Approche 2 : Spectrogramme du signal bruité | 64 |
| 4.9 | Approche 2 : Spectrogramme du bruit estimé avec VAD énergétique | 64 |
| 4.10 | Approche 2 : Spectrogramme du signal nettoyé avec VAD énergétique | 64 |
| 4.11 | Approche 3 : Spectrogramme du signal bruité | 66 |
| 4.12 | Approche 3 : Spectrogramme du bruit estimé avec VAD basé sur SVM | 67 |
| 4.13 | Approche 3 : Spectrogramme du signal nettoyé avec VAD basé sur SVM | 67 |
| 4.14 | Matrice de confusion du détecteur VAD basé sur SVM. | 68 |

Liste des tableaux

| | | |
|-----|---|----|
| 4.1 | Résumé des fonctions principales du traitement audio | 60 |
| 4.2 | Avantages et limites de la soustraction spectrale avec VAD statique | 63 |
| 4.3 | Avantages et limites de la soustraction spectrale avec VAD énergétique | 65 |
| 4.4 | Caractéristiques utilisées pour la classification VAD. | 66 |
| 4.5 | Configuration de l'algorithme SVM pour le VAD. | 66 |
| 4.6 | Rapport de classification du VAD-SVM. | 67 |
| 4.7 | Avantages et limites de la soustraction spectrale avec VAD basé sur SVM | 68 |
| 4.8 | Performances comparées selon le critère SDR. | 69 |

Introduction Générale

La compréhension et la restitution de la parole en environnement bruité représentent un enjeu majeur dans le domaine de l'assistance auditive. Les personnes malentendantes sont particulièrement pénalisées dans ces contextes, où le signal vocal se trouve souvent masqué par des interférences sonores ambiantes. Pour améliorer leur confort d'écoute, les aides auditives modernes doivent intégrer des mécanismes efficaces de réduction de bruit.

C'est dans ce cadre que s'inscrit le présent travail, qui combine des outils classiques du traitement du signal avec des techniques d'intelligence artificielle. L'analyse fréquentielle, et en particulier la **transformée de Fourier**, joue un rôle central dans ce contexte. En permettant de représenter un signal comme une superposition de composantes sinusoïdales, la transformée de Fourier offre un accès direct au contenu spectral du signal. Cette représentation rend possible la séparation entre les composantes vocales (structurelles, harmoniques) et le bruit (souvent aléatoire et réparti spectralement).

Parmi les techniques exploitant cette transformation, la **soustraction spectrale** est l'une des méthodes les plus utilisées en réduction de bruit. Elle consiste à estimer le spectre du bruit, puis à le soustraire du spectre du signal bruité. Toutefois, cette méthode reste limitée dans des environnements non stationnaires, notamment à cause d'une détection imprécise des segments non vocaux servant à l'estimation du bruit.

C'est pour pallier ces lacunes qu'une approche hybride est proposée dans ce travail. Elle consiste à coupler la soustraction spectrale avec un modèle d'apprentissage supervisé : la *machine à vecteurs de support* (SVM). Ce modèle permet de détecter plus précisément les instants sans parole, améliorant ainsi la qualité de l'estimation du bruit et, par conséquent, la performance globale de la réduction.

L'objectif de ce mémoire est donc de concevoir, implémenter et comparer trois variantes de la soustraction spectrale : une version basique avec estimation statique, une version améliorée avec un VAD (Voice Activity Detection) énergétique, et une version intelligente intégrant un VAD basé sur SVM. Les performances seront évaluées à l'aide du critère SDR (*Signal-to-Distortion Ratio*) et les spectrogrammes du signal débruité.

Le mémoire est structuré en quatre chapitres complémentaires, que l'on peut brièvement décrire comme suit :

Le **chapitre 1** : introduit les bases physiologiques de l'audition en décrivant l'anatomie de l'oreille humaine, les principales pathologies auditives, notamment la surdité, ainsi que le rôle et le fonctionnement des aides auditives.

Le **chapitre 2** : est consacré aux fondements du traitement des signaux audio. Il présente la nature physique du son, les différentes représentations des signaux (temporelle et fréquentielle), les caractéristiques du

bruit, ainsi que les indicateurs de qualité utilisés pour évaluer les performances des systèmes de traitement.

Le **chapitre 3** : présente la transformée de Fourier et ses variantes numériques, en particulier la transformée discrète (TFD), la transformée rapide (TFR) et la transformée à court terme (TFCT). Il met en évidence le rôle fondamental de la TFCT dans l'analyse spectrale locale, ainsi que son utilisation dans plusieurs techniques de réduction du bruit.

Enfin, le **chapitre 4** : est dédié à l'étude de la méthode de soustraction spectrale. Il en détaille les fondements théoriques, les principales variantes (statique, adaptative, probabiliste), ainsi que les limites rencontrées en pratique. Trois approches d'implémentation sont comparées, incluant une méthode hybride intégrant un classifieur SVM pour améliorer la détection de la parole et l'efficacité globale de la réduction de bruit dans les environnements complexes.

Chapitre 1

L'oreille humaine : Structure, pathologies et solutions technologiques

1.1 Introduction

Ce chapitre introduit les bases de l'audition humaine en explorant d'abord les propriétés physiques du son, comme la fréquence et l'amplitude, qui influencent la perception auditive. Il détaille ensuite l'anatomie de l'oreille externe, moyenne et interne, et leur rôle dans la transformation du son en signaux nerveux. Une attention particulière est portée aux pathologies auditives, notamment la surdité sous ses différentes formes et causes. Les conséquences sur la qualité de vie, telles que les difficultés de communication et les acouphènes, sont également abordées. Le chapitre se termine par une présentation des aides auditives, expliquant leur fonctionnement et leur importance dans l'amélioration de l'autonomie des personnes malentendantes.

1.2 Anatomie de l'oreille humaine

L'oreille est un organe complexe qui transforme les ondes sonores en signaux électriques interprétés par le cerveau. Elle est divisée en trois parties principales voir[14] : l'oreille externe, l'oreille moyenne et l'oreille interne. Comme le montre la figure 1.1.

L'oreille externe

L'oreille externe, véritable porte d'entrée du système auditif, assure la captation et la transmission initiale des ondes sonores aériennes vers l'oreille moyenne. Elle se compose de deux éléments essentiels :

- Le **pavillon**, par sa forme, collecte les ondes sonores et les dirige efficacement vers le conduit auditif, jouant également un rôle crucial dans la localisation des sources sonores.
- Le **conduit auditif**, agit comme un canal amplificateur, ciblant spécifiquement les fréquences situées entre 2 et 5 kHz, une plage essentielle pour la perception de la parole. Il achemine ensuite les ondes sonores amplifiées jusqu'au tympan, marquant ainsi la transition vers l'oreille moyenne.

L'oreille moyenne

L'oreille moyenne, élément central de la transmission sonore, est constituée de trois composants essentiels :

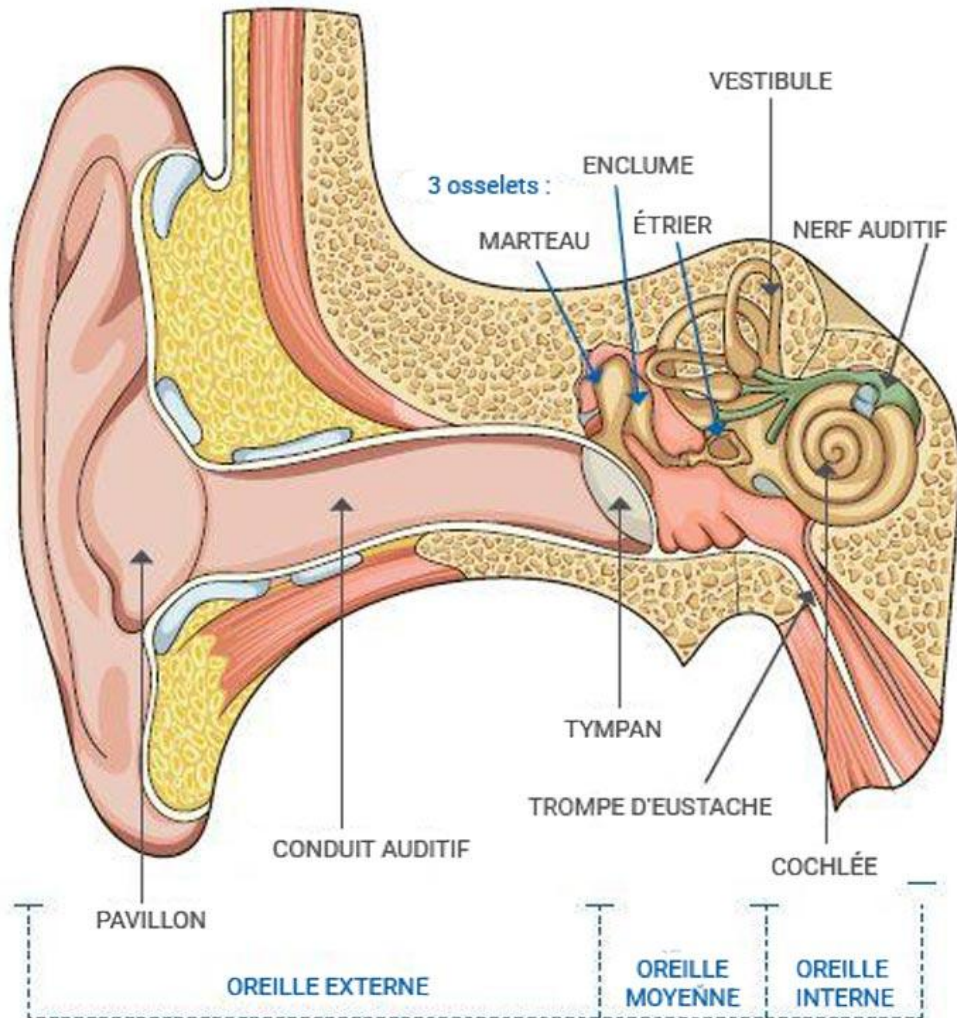


FIGURE 1.1 – Anatomie de l'oreille

- La **membrane tympanique**, qui vibre sous l'effet des ondes.
- La **chaîne des osselets** composée du marteau, de l'enclume et de l'étrier, qui amplifie mécaniquement ces vibrations et les transmet à l'oreille interne via la fenêtre ovale.
- La **trompe d'Eustache**, équilibre la pression entre l'oreille moyenne et l'environnement extérieur.

L'oreille interne

L'oreille interne est l'organe principal de l'audition. Elle transforme les signaux acoustiques en messages nerveux grâce à deux structures clés :

- La **cochlée**, responsable de l'audition
- Le **vestibule**, dédié à l'équilibre

Au sein de la cochlée, l'**organe de Corti** repose sur la membrane basilaire et contient :

- Les **cellules ciliées internes (CCI)** : elles transmettent les signaux au nerf auditif (la figure 1.2).
- Les **cellules ciliées externes (CCE)** : elles amplifient activement les vibrations grâce à l'électromotilité.

Leurs stéréocils réagissent aux mouvements du liquide cochléaire, permettant la détection de sons extrêmement faibles (seuil < 0 dB SPL).

Remarque 1 La cochlée, illustrée par la figure 1.3 véritable cœur de la transduction auditive, joue un rôle

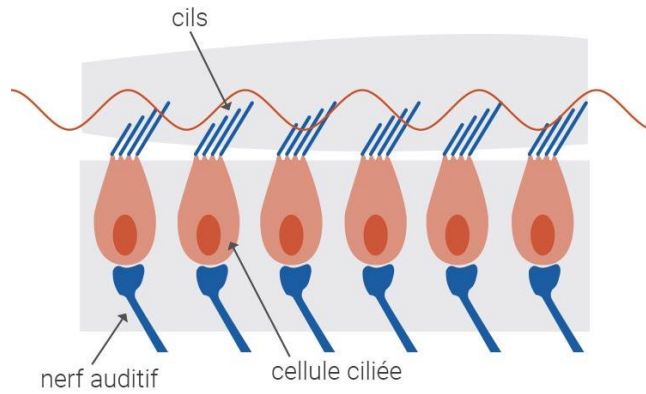


FIGURE 1.2 – Cellules ciliées

fondamental dans notre capacité à entendre. Organe spiralé situé dans l'os temporal, elle transforme les ondes sonores en impulsions nerveuses grâce à une structure complexe : divisée en rampes vestibulaire et tympanique, elle abrite le canal cochléaire contenant l'organe de Corti, sur la membrane basillaire. Les vibrations de l'étrier, via la fenêtre ovale, se propagent dans le liquide cochléaire, stimulant les cellules ciliées et initiant ainsi le processus de l'audition, avant d'être évacuées par la fenêtre ronde.

Le rôle de la cochlée

Organe spiralé de l'os temporal, la cochlée assure la transduction des ondes sonores en impulsions nerveuses.

Sa structure comprend :

- Deux rampes liquidiennes (vestibulaire et tympanique)
- Le canal cochléaire, contenant l'organe de Corti.

Les vibrations transmises par l'étrier via la fenêtre ovale propagent des ondes dans le liquide cochléaire. Ce mouvement stimule les cellules ciliées de la membrane basillaire, initiant la transduction auditive, avant que l'énergie résiduelle ne soit évacuée par la fenêtre ronde.

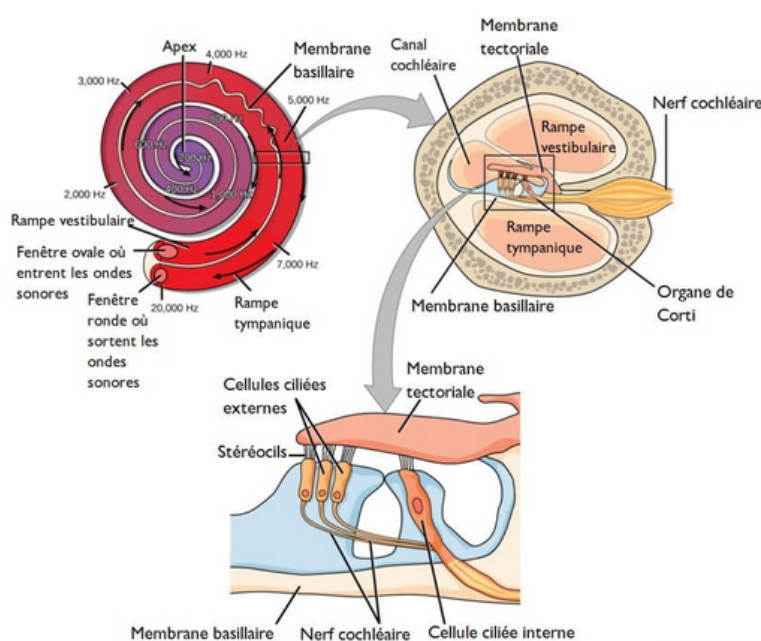


FIGURE 1.3 – La cochlée

1.3 Les maladies de l'oreille

L'oreille humaine est un système sensoriel exceptionnel, capable de détecter une large gamme de sons, des plus faibles aux plus forts. Sa sensibilité est remarquable, malgré un nombre limité de cellules sensorielles. Cependant, cette complexité rend l'oreille vulnérable à diverses maladies, pouvant entraîner des troubles auditifs.[2] Cette vulnérabilité se manifeste par une multitude d'affections susceptibles d'affecter toutes les parties de l'oreille, ce qui peut entraîner non seulement des troubles auditifs, mais aussi des complications de santé plus larges.

Oreille Externe

L'oreille externe, qui se compose du pavillon et du conduit auditif, peut être affectée par diverses maladies.

— Pavillon de l'oreille

Le pavillon de l'oreille peut être sujet à plusieurs affections. L'une d'elles est la *chondrite auriculaire*, une inflammation du cartilage de l'oreille souvent causée par une infection bactérienne. De plus, des kystes épidermiques ou des tumeurs bénignes peuvent également se développer sur le pavillon.

— Conduit auditif

Le conduit auditif peut souffrir de plusieurs problèmes de santé. L'*otite externe*, par exemple, est une infection souvent causée par l'eau. Par ailleurs, une accumulation excessive de cérumen peut entraîner une perte auditive temporaire. Enfin, la *dermatite séborrhéique* peut provoquer des démangeaisons et des irritations dans cette zone.

Oreille moyenne

L'oreille moyenne, qui se trouve entre le tympan et l'oreille interne, peut être touchée par plusieurs pathologies :

— Tympan

Le tympan peut être affecté par plusieurs problèmes, notamment la perforation du tympan, qui est une rupture de la membrane tympanique souvent causée par une infection ou un traumatisme. De plus, l'otite moyenne est une infection de l'oreille moyenne, pouvant être aiguë ou chronique, et est souvent liée à des infections respiratoires.

— Chaîne des osselets

La chaîne des osselets, qui joue un rôle crucial dans la transmission du son, peut être affectée par plusieurs conditions. L'otospongiose est une maladie héréditaire qui impacte la mobilité de l'étrier, entraînant une perte auditive de transmission. De plus, une dislocation ou une fracture des osselets peut survenir à la suite d'un traumatisme crânien, ce qui peut également provoquer une perte auditive.

— Trompe d'Eustache

La trompe d'Eustache peut présenter des dysfonctionnements qui entraînent des problèmes d'équilibre de pression, des douleurs, ainsi que des infections de l'oreille moyenne. De plus, une obstruction de cette trompe peut provoquer une otite moyenne avec effusion, caractérisée par une accumulation de liquide dans l'oreille moyenne.

Oreille Interne

L'oreille interne peut également être touchée par diverses maladies qui affectent l'audition et l'équilibre

- **Cochlée** La cochlée, une structure de l'oreille interne, est cruciale pour l'audition. La surdité neurosensorielle résulte de dommages aux cellules ciliées de la cochlée, souvent causés par des traumatismes sonores ou le vieillissement. Par ailleurs, la maladie de Ménière, également liée à la cochlée, se manifeste par des vertiges, des acouphènes et une perte auditive fluctuante. Ces conditions soulignent l'importance de la cochlée pour l'audition et l'équilibre.
- **Organe de Corti** L'organe de Corti, essentiel à l'audition, peut subir des dommages aux cellules ciliées, entraînant une perte auditive souvent causée par des bruits forts ou des facteurs génétiques. Ces dommages peuvent également provoquer des acouphènes, des sons perçus sans source externe. Il est donc crucial de protéger cet organe pour maintenir une bonne audition.

1.3.1 La surdité

Parmi les troubles les plus significatives de l'oreille, la surdité constitue une des pathologies du système auditif. Elle se définit comme une perte partielle ou totale de l'audition. La classification de la surdité repose sur l'identification de la zone de l'oreille atteinte. Ainsi, nous distinguons la surdité de transmission, de perception, et la surdité mixte (la figure 1.4).

- **La surdité de transmission :** Elle résulte d'une atteinte de l'oreille externe ou moyenne, affectant des éléments tels que le conduit auditif, le tympan ou la chaîne des osselets. Ce type de surdité peut être causé par une mauvaise hygiène de l'oreille, entraînant la formation de bouchons de cérumen, mais la cause la plus courante est une infection de l'oreille, connue sous le nom d'otite. La surdité de transmission se manifeste par une atténuation de l'onde acoustique, qui peut varier de 20 à 60 dB.
- **La surdité de perception :** La surdité de perception, qui affecte l'oreille interne, résulte d'une atteinte des cellules ciliées ou des nerfs auditifs. Cette atteinte perturbe la transmission des messages nerveux, qui sont soit atténués, soit mal interprétés par le cerveau.
- **La surdité mixte :** La surdité mixte, combinant les atteintes de l'oreille moyenne ou externe et interne, résulte souvent d'infections et peut endommager le tympan, les osselets et les liquides internes ; parallèlement, la presbyacousie, une forme de surdité de perception liée à l'âge, affecte progressivement les cellules ciliées de la cochlée, surtout dans les hautes fréquences, et devient de plus en plus courante avec l'allongement de l'espérance de vie. Pour diagnostiquer ces types de surdité, l'audiométrie tonale est essentielle, mesurant les pertes auditives à diverses fréquences (125 Hz à 8000 Hz) et permettant de calculer la perte tonale moyenne, principalement à partir des fréquences de 500 Hz, 1000 Hz, 2000 Hz et 4000 Hz.[14]

La surdité peut être le résultat de multiples facteurs, allant de l'usure naturelle liée au vieillissement à l'exposition répétée à des environnements sonores intenses, en passant par des prédispositions génétiques ou des complications infectieuses.

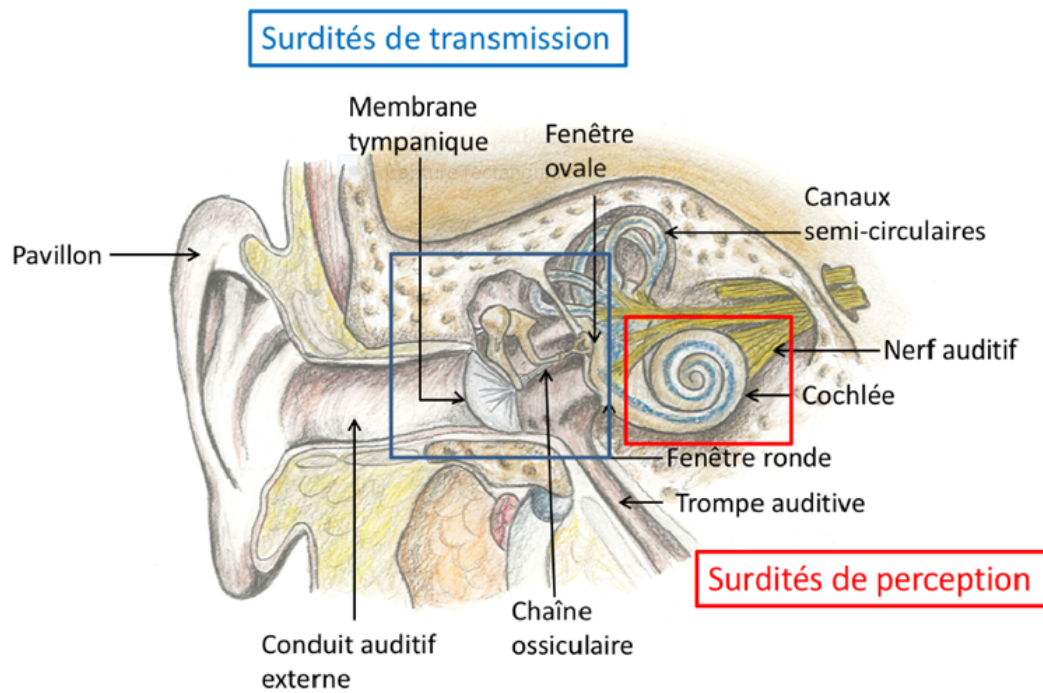


FIGURE 1.4 – La surdité

Face à la diversité des maladies de l'oreille et à leurs impacts potentiels sur l'audition, il est crucial de considérer les solutions disponibles pour atténuer ces troubles. Les prothèses auditives représentent une option efficace pour les personnes souffrant de pertes auditives, qu'elles soient causées par des affections temporaires ou chroniques. Ces dispositifs permettent non seulement d'améliorer la qualité de vie des utilisateurs, mais aussi de leur redonner une certaine autonomie dans leur communication quotidienne.

1.4 Prothèses auditives

L'appareil auditif, également connu sous le nom de prothèse auditive ou encore aide auditive (la figure 1.5), est un dispositif électronique conçu pour aider les personnes souffrant de perte auditive à percevoir les sons. Il fonctionne en amplifiant les sons ambiants afin de les rendre plus audibles pour l'utilisateur. [18]

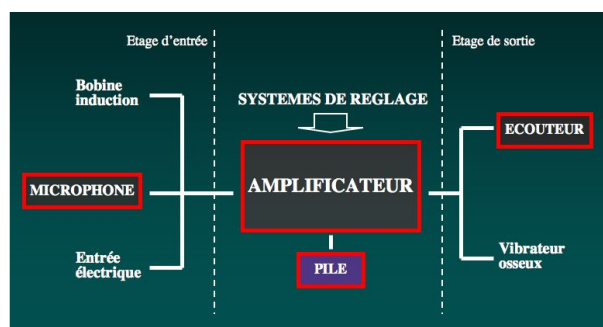


FIGURE 1.5 – Schéma de l'appareil auditif

1.4.1 Les composants principaux d'un appareil auditif

Les appareils auditifs fonctionnent en captant, amplifiant et restituant les sons. Ils sont constitués de plusieurs composants importants (la figure 1.8).

- **Le processeur**

Le processeur dans une aide auditive est le composant central qui traite les sons captés par le microphone. Il agit comme le "cerveau" de l'appareil.

- **Microphone** Le microphone de l'appareil auditif est un composant essentiel qui capte les sons environnants. Il transforme les ondes sonores, qui sont des signaux acoustiques, en signaux électriques. Ces signaux sont ensuite convertis en signaux numériques, permettant ainsi une amplification adaptée aux besoins auditifs de l'utilisateur. Deux types de microphones sont utilisés dans le domaine de l'audioprothèse.

- **Microphones omnidirectionnels**

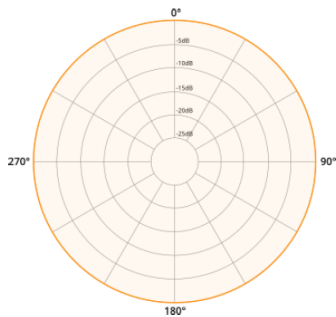


FIGURE 1.6 – Microphones omnidirectionnels

Les microphones omnidirectionnels, illustrés par la figure 1.6, se distinguent par leur conception simple et leur capacité à capter les sons de manière uniforme à 360 degrés, ce qui en fait des références pour d'autres types de microphones, cependant, leur performance peut varier en hautes fréquences, où ils tendent à devenir directionnels, et cette omnidirectionnalité est directement influencée par leur taille, les modèles plus petits conservant mieux leur comportement omnidirectionnel en raison de la réduction de l'effet de blocage des ondes sonores par le boîtier.[5] Les microphones omnidirectionnels se caractérisent par leur sensibilité égale à toutes les sources sonores environnantes, comme en témoigne leur diagramme polaire circulaire à 360 degrés.

- **Microphones directionnels**

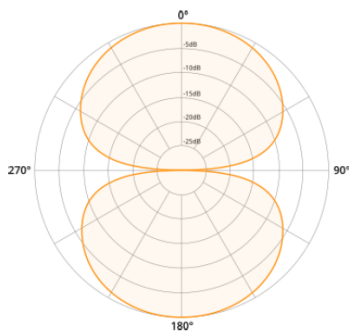


FIGURE 1.7 – Microphones directionnels (bidirectionnels)

Les microphones directionnels, illustrés par la figure 1.7, sont spécialement conçus pour la prise de son à l'avant (et également à l'arrière dans le cas des modèles bidirectionnels), en éliminant le son hors axe. Cet effet varie selon les fréquences et seuls les meilleurs modèles garantissent une élimination uniforme dans une large bande de fréquences. La directivité dépend des ouvertures externes et passages internes du microphone permettant au son d'atteindre précisément les deux côtés de la membrane. Le son parvenant de l'avant du microphone fera vibrer la membrane, alors que celui de l'arrière annulera cette vibration.[5]

- **Amplificateur**

L'amplificateur, qui est un élément important de la prothèse auditive, modifie et amplifie le signal électrique provenant du microphone. Ce signal est numérisé à l'aide d'algorithmes avancés et traité selon divers paramètres. Deux systèmes d'amplification sont généralement utilisés : l'amplification linéaire, qui

augmente tous les sons de manière uniforme, et l'amplification non linéaire, qui ajuste l'amplification en fonction de la perte auditive à certaines fréquences.

- **L'écouteur**

Il convertit le signal électrique, modifié par l'amplificateur, en signal acoustique afin de diriger les sons vers le tympan. Ainsi, il transmet le son traité par l'amplificateur vers le conduit auditif.[6]

- **Source d'énergie**

Les appareils sont alimentés par des piles de différentes tailles, adaptées à des besoins et des consommations variés en fonction de la puissance et du type d'appareil. Récemment, certains fabricants ont commencé à proposer des appareils rechargeables, qui peuvent fonctionner avec des piles « classiques » ou des accumulateurs, offrant ainsi une solution beaucoup plus pratique pour l'utilisateur .



FIGURE 1.8 – Les composants principaux d'un appareil auditif

1.4.2 Fonctionnement technique d'une aide auditive

L'aide auditive fonctionne en plusieurs étapes : tout d'abord, les microphones captent les sons de l'environnement et transforment l'énergie acoustique en signal électrique. Ce signal est ensuite envoyé à l'amplificateur, souvent intégré dans le processeur de l'appareil, qui amplifie et modifie le signal électrique en fonction des besoins auditifs spécifiques de l'utilisateur, tenant compte de sa perte auditive. Le signal amplifié est alors transmis à l'écouteur, qui convertit l'énergie électrique en énergie acoustique, restituant ainsi le son à l'intérieur de l'oreille de l'utilisateur, ce qui permet une meilleure perception des sons environnants. De plus, l'utilisateur peut ajuster le volume à l'aide d'un potentiomètre, personnalisant ainsi son expérience auditive en fonction de l'environnement sonore.[17]

1.5 Conclusion

Ce chapitre a exploré l'anatomie et le rôle fondamental de l'oreille humaine dans la perception auditive, depuis la capture du son par l'oreille externe jusqu'à sa conversion en signaux électriques dans l'oreille interne. Il a aussi mis en lumière les principales infections auditives telles que la presbycusie, les traumatismes sonores. Pour y remédier, les technologies comme les appareils auditifs et les implants cochléaires jouent un rôle essentiel.

Ces dispositifs, bien qu'imparfaits, compensent partiellement les pertes auditives. Leur évolution découle d'une meilleure compréhension des mécanismes de l'audition. Ils contribuent ainsi à restaurer la communication et à améliorer la qualité de vie des personnes malentendantes.

Chapitre 2

Fondements du traitement de signal

2.1 Introduction

Ce chapitre aborde les fondements du traitement du signal, en mettant l'accent sur les concepts essentiels de l'analyse et de la manipulation des signaux sonores. Nous commencerons par définir le signal sonore, en explorant sa nature et ses propriétés physiques. Ensuite, nous examinerons les différentes représentations et classifications des signaux, ainsi que la nature et l'impact du bruit sur la perception auditive. Enfin, nous discuterons des principes de réduction du bruit et des techniques disponibles pour améliorer la qualité des signaux audio, en introduisant les concepts de **rapport signal à bruit (RSB)** et le **rapport signal sur distorsion (SDR)** comme mesures clés de performance des appareils auditifs. Ce chapitre permettra de mieux comprendre les défis et les solutions associés au traitement des signaux sonores.

2.2 Définitions et exemples

Un signal peut être défini comme une fonction mathématique représentant l'évolution d'une grandeur physique (tension, intensité, pression, température, etc.) en fonction d'une ou plusieurs variables indépendantes, souvent le temps ou l'espace.

Les signaux sont omniprésents dans de nombreux domaines scientifiques et techniques, on peut citer :

- L'acoustique : les sons perçus par l'oreille sont des variations de pression de l'air.
- L'électronique : les signaux électriques (courant, tension) évoluant dans les circuits.
- Traitement d'images : une image peut être vue comme un signal bidimensionnel en fonction de deux variables spatiales.
- Médecine : les électrocardiogrammes (ECG), électroencéphalogrammes (EEG), etc.

Formulation mathématique d'un signal

$$x : D \subseteq \mathbb{R}^n \rightarrow \mathbb{K}, \quad \mathbb{K} = \mathbb{R} \text{ ou } \mathbb{C},$$

Dans le cas des *signaux audio*, comme ceux analysés dans les appareils auditifs, on considère généralement

un signal temporel :

$$x : \mathbb{R} \rightarrow \mathbb{R}, \quad t \mapsto x(t),$$

où $x(t)$ désigne l'amplitude de l'onde sonore au temps t . Ces signaux sont souvent supposés à énergie finie, $x \in L^2(\mathbb{R})$, c'est-à-dire des fonctions à valeurs réelles, mesurables et à carré intégrable :

$$\int_{\mathbb{R}} |x(t)|^2 dt < \infty.$$

Ce cadre permet l'utilisation d'outils analytiques puissants comme la transformée de Fourier, la transformée en ondelettes, ainsi que de méthodes algorithmiques modernes : Filtrage adaptatif, réseaux de neurones profonds.

2.2.1 Le signal sonore : Nature et propriétés physiques

Définition 1 *Le son est une onde mécanique qui se propage à travers un milieu élastique, comme l'air, les liquides ou les solides. Il est produit par les vibrations d'une source, qui génère des variations de pression dans le milieu environnant (la figure 2.1).[13]*

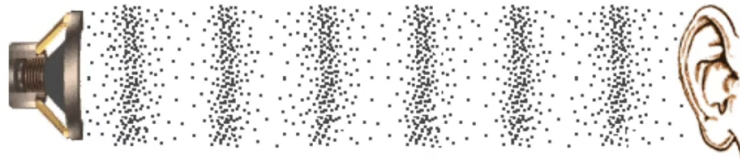


FIGURE 2.1 – Variation de la pression de l'air dans le temps

Caractéristiques physiques du son : Le son, est défini par sa fréquence, son amplitude, sa longueur d'onde, sa vitesse de propagation, et son spectre. Ces paramètres permettent de caractériser et de différencier les sons entre eux.

Amplitude : C'est une mesure de l'intensité du son. Elle influe sur la perception du volume sonore : Une grande amplitude produit un son fort, une petite amplitude produit un son faible (la figure 2.2).

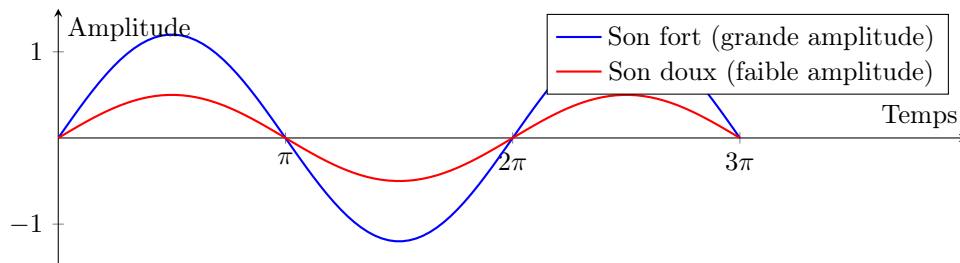


FIGURE 2.2 – Représentation de l'amplitude d'un signal sonore

Fréquence : Généralement notée par f , elle désigne le nombre d'oscillations par seconde, mesuré en Hertz (Hz). Elle détermine la hauteur du son : Les fréquences basses sont perçues comme des sons graves, et les hautes comme des sons aigus (figure 2.3).

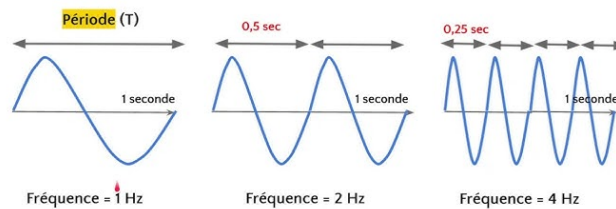


FIGURE 2.3 – La fréquence

Longueur d'onde (λ) : La distance entre deux points correspondants successifs d'une onde (par exemple, deux crêtes). Relation avec la fréquence : $\lambda = \frac{v}{f}$, où v est la vitesse du son.

Vitesse de propagation (v) : La vitesse à laquelle une onde se déplace à travers un milieu :

- Dans l'air à 20°C : Environ 343 m/s.
- Plus grande dans les liquides et les solides.

Spectre du son : Le spectre d'un son, ou d'une onde en général, illustré par la figure 2.4, montre la composition de ce son : En abscisse, les différentes fréquences qui le composent, et en axe des ordonnées, l'amplitude de chaque raie. Le spectre est obtenu à l'aide de l'algorithme TFR, l'outil mathématique qui nous permet de passer du domaine temporel (signal) au domaine fréquentiel (spectre) tel que :

- Les sons purs contiennent une seule fréquence (ex. : Diapason).
- Les sons complexes, comme la voix ou la musique, sont composés de multiples fréquences.

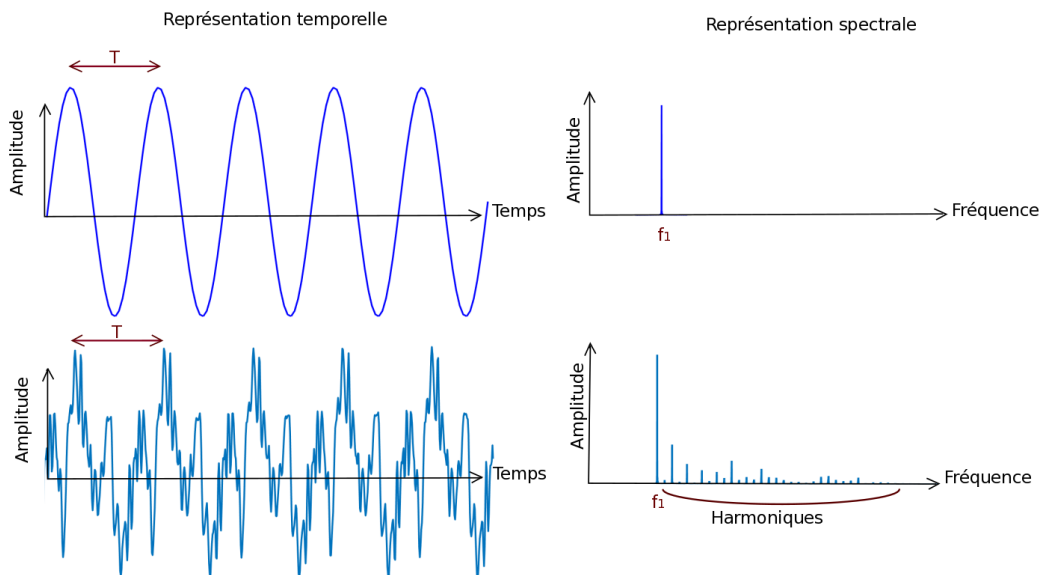


FIGURE 2.4 – Exemples de spectres du son

Traitement du signal : Dans le domaine du traitement du signal, un signal véhicule une information qui peut couvrir différentes formes selon le contexte et l'application considérée. Cette information est souvent cachée dans les variations du signal au cours du temps ou de l'espace, et l'objectif du traitement du signal est de la mettre en évidence, l'extraire, la modifier ou la transmettre de manière optimale.

Par exemple, dans le cas d'un signal audio, l'information peut correspondre à une parole ou une musique que l'on souhaite rendre plus claire ou compréhensible, notamment en réduisant le bruit de fond. Dans le domaine médical, les signaux issus de capteurs (comme un électrocardiogramme) contiennent des informations cruciales qui doivent être analysées avec précision pour diagnostiquer ou surveiller un état pathologique. Le traitement du signal englobe ainsi des opérations diverses telles que :

- **La filtration** ; pour supprimer les composantes indésirables (bruit, interférences) ;
- **La transformation** ; pour changer la représentation du signal (exemple via la transformée de Fourier) ;
- **La compression** ; pour réduire la taille des données tout en conservant l'information essentielle ;
- **La reconstruction** ; pour restituer un signal à partir de mesures incomplètes ou bruitées ;
- **La classification ou la reconnaissance** ; pour identifier des motifs ou des caractéristiques spécifiques dans le signal.

Ainsi, la richesse et la complexité d'un signal imposent le développement de méthodes adaptées, combinant des outils mathématiques, algorithmiques et informatiques, afin d'exploiter pleinement l'information qu'il contient.

2.3 Représentation et classification des signaux

Cette section présente les différentes manières de représenter les signaux ainsi que leur classification selon leurs caractéristiques.

2.3.1 Représentation des signaux

La représentation des signaux peut être envisagée selon deux approches complémentaires :

- **Représentation temporelle** : La représentation temporelle d'un signal décrit l'évolution de son amplitude au cours du temps. Elle permet d'observer directement les variations instantanées, les changements de forme d'onde ainsi que les fluctuations d'intensité. Cette représentation est couramment utilisée dans l'analyse des signaux audio, vidéo ou électriques, sous forme de graphiques où l'axe horizontal représente le temps et l'axe vertical représente l'amplitude.
- **Représentation fréquentielle ou spectrale** : La représentation fréquentielle met en évidence les composantes en fréquence d'un signal. Elle permet de décomposer celui-ci en une somme de sinusoides de différentes fréquences, amplitudes et phases. Cette représentation, souvent illustrée par un spectre de fréquences, révèle les fréquences présentes dans le signal ainsi que leur intensité. Elle constitue un outil essentiel pour l'analyse des signaux, la conception de filtres et l'étude des systèmes linéaires.

2.3.2 Classification des signaux

Dans cette section, nous exposons des critères de classification essentiels pour bien appréhender la variété des signaux et adapter les méthodes d'analyse et de traitement appropriées.

En effet, les signaux présentent une grande diversité qui influe directement sur leur analyse et leur traitement. Pour cela, plusieurs critères de classification sont utilisés afin de regrouper les signaux selon leurs caractéristiques fondamentales (continuité, périodicité, nature déterministe ou aléatoire, stationnarité, etc.). Les signaux peuvent être classifiés selon plusieurs critères :

1. Classification selon la continuité

- **Signaux analogiques** : Un signal analogique (la figure 2.5) est un signal défini en temps continu, dont la grandeur (amplitude) peut prendre n'importe quelle valeur dans un intervalle donné. Il est représenté par une fonction continue du temps, c'est-à-dire que sa valeur est disponible à tout instant. Les signaux analogiques sont généralement représentés dans un plan à deux dimensions, où l'axe des abscisses représente le temps et l'axe des ordonnées l'amplitude du signal.

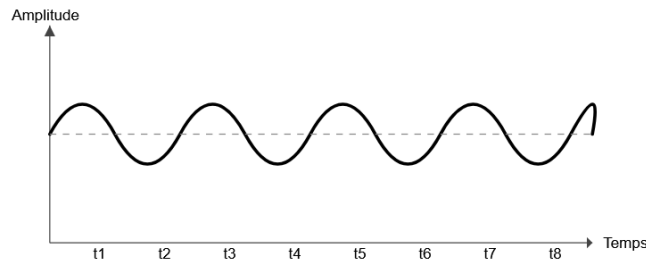


FIGURE 2.5 – Signal analogique

Exemple 2.3.1

1. **Signal triangle** : Un signal triangulaire symétrique centré à l'origine peut être défini comme suit (la figure 2.6) :

$$x(t) = \begin{cases} \frac{1}{\epsilon} \left(1 - \frac{|t|}{\epsilon}\right) & \text{si } |t| \leq \epsilon \\ 0 & \text{sinon} \end{cases}$$

où ϵ est la largeur du triangle.

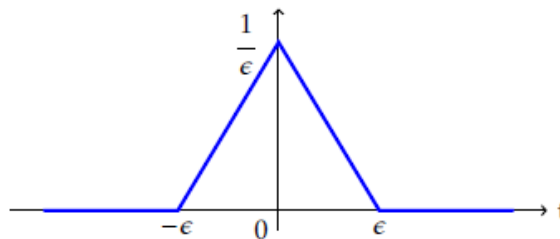


FIGURE 2.6 – Signal triangle

2. Impulsion de Dirac $\delta(t)$

Elle est illustrée par la figure 2.7. Elle est définie comme limite d'une fonction triangulaire

$$\delta(t) = \lim_{\epsilon \rightarrow 0} x_{\epsilon}(t)$$

La fonction δ vérifie

$\delta(t) = 0$ pour $t \neq 0$. et

$$\int_{-\infty}^{+\infty} \delta(t) dt = 1.$$

Lorsqu'une impulsion se produit au temps $t = a$, on la note $\delta(t - a)$.

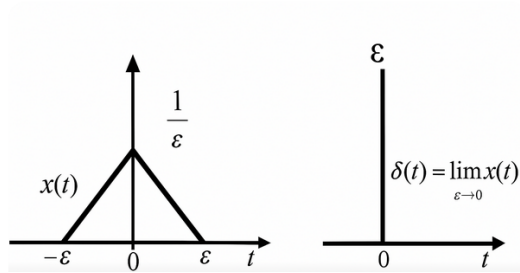


FIGURE 2.7 – Impulsion de Dirac

La fonction impulsion est fréquemment rencontrée dans l'étude des signaux, notamment lors de commutations rapides ou lorsque des circuits sont excités par des sources impulsionnelles. Elle constitue également un outil mathématique fondamental.

Lorsque ε tend vers 0, les propriétés suivantes s'observent :

- (a) L'amplitude tend vers l'infini.
- (b) La durée du pulse diminue vers zéro.
- (c) L'aire totale du triangle reste constante et égale à 1.

La figure suivante montre que lorsque ε tend vers 0, la fonction triangulaire tend vers l'impulsion de Dirac

- **Signaux numériques à temps discret**

On dit qu'un signal est **numérique à temps discret** (la figure 2.8) lorsqu'il est observé uniquement à des instants discrets, généralement espacés de manière régulière. Formellement, il est défini comme suit :

Un **signal discret** est une suite d'échantillons, c'est-à-dire une fonction définie sur \mathbb{Z} (ou une sous-partie de \mathbb{Z}) :

$$x : \mathbb{Z} \rightarrow \mathbb{R} \quad \text{ou} \quad x : \mathbb{Z} \rightarrow \mathbb{C}$$

En pratique, on considère souvent un signal de durée finie, défini pour $n \in \{0, 1, \dots, N - 1\}$.

La notation $x[n]$ désigne la valeur du signal à l'instant n , c'est-à-dire l'échantillon numéro n . Par extension, on utilise aussi $x[n]$ pour désigner l'ensemble du signal, de la même manière que dans le cas continu où $x(t)$ désigne tout le signal x .

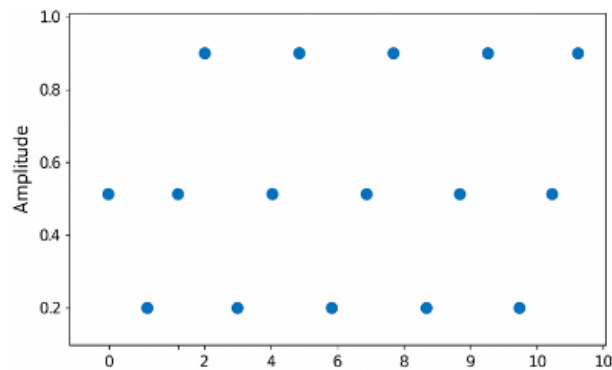


FIGURE 2.8 – Signal numérique

Remarque 2 Pour permettre le traitement d'un signal temporel $x(t)$ à l'aide d'un ordinateur numérique (comme un micro-ordinateur), il est nécessaire de procéder à une numérisation.

Numérisation d'un signal analogique : La **numérisation** d'un signal analogique est le processus qui permet de transformer un signal continu en un signal numérique, exploitable par un système informatique. Ce processus se déroule en trois étapes fondamentales : *l'échantillonnage, la quantification, et le codage.*

(a) **Échantillonnage**

L'échantillonnage est le processus de conversion d'un signal analogique continu en une suite discrète de valeurs prises à intervalles réguliers.[20, 19]

Grâce à l'échantillonnage, il devient possible d'appliquer des outils numériques comme la Transformée de Fourier Discrète (TFD) et son algorithme efficace, la TFR qui permettent d'analyser le contenu fréquentiel du signal échantillonné.

L'échantillonnage d'un signal analogique consiste à prélever des valeurs de ce signal à des instants précis, espacés régulièrement dans le temps.

Un objectif fondamental de l'échantillonnage est de préserver l'information du signal analogique. Cela signifie qu'il doit être possible de reconstituer fidèlement le signal original à partir de ses échantillons Pour garantir cette possibilité de reconstitution fidèle du signal original, il est essentiel de comprendre les conditions mathématiques qui rendent cela possible. C'est précisément ce que formalise le théorème de Nyquist-Shannon ^{1 2}.

Théorème 1 (Shannon, 1949) [19]

Soit $x(t)$ un signal continu bande limitée à la fréquence f_{max} , c'est-à-dire que sa transformée de Fourier $X(f)$ est nulle ou négligeable pour $|f| > f_{max}$. Alors, $x(t)$ est entièrement déterminé par ses échantillons pris aux instants $t_n = \frac{n}{2f_{max}}$, avec $n \in \mathbb{Z}$, et il peut être reconstruit exactement par :

$$x(t) = \sum_{n=-\infty}^{+\infty} x\left(\frac{n}{2f_{max}}\right) \cdot \text{sinc}\left(2f_{max}\left(t - \frac{n}{2W}\right)\right)$$

où $W = \frac{n}{2f_{max}}$

et la fonction sinus cardinal "sinc" est définie par :

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$

Ainsi, afin de respecter les conditions du théorème, l'échantillonnage pratique s'effectue selon les deux étapes suivantes :

i. **Choix de la fréquence d'échantillonnage f_e**

Selon le Théorème de Nyquist-Shannon il faut choisir une fréquence d'échantillonnage f_e au moins égale au double de sa fréquence maximale, $f_e \geq 2f_{max}$. Cette fréquence est obtenue en identifiant la bande passante du signal.

ii. **Discretisation temporelle**

1. Harry Nyquist (1889–1976), Ingénieur et physicien suédo-américain.

Il est à l'origine du concept de taux minimum d'échantillonnage pour éviter le repliement spectral (aliasing)

2. Claude Shannon (1916–2001) Mathématicien, ingénieur et père de la théorie de l'information. En 1949, il formalise dans un cadre mathématique rigoureux le théorème d'échantillonnage, désormais connu sous le nom de théorème de Nyquist-Shannon.

Une fois la fréquence d'échantillonnage déterminée, on prend $T_e = \frac{1}{f_e}$ période d'échantillonnage, on procède ensuite à la discrétisation proprement dite du signal dans le domaine temporel. On prélève la valeur du signal à chaque instant $t_n = nT_e$, instants d'échantillonnage, où n est un entier relatif (la figure 2.9).

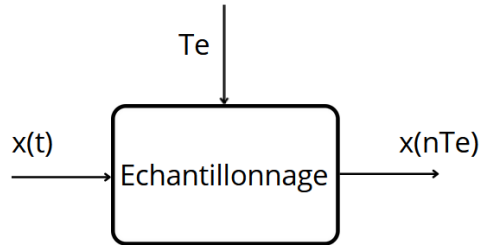


FIGURE 2.9 – Schéma d'échantillonnage d'un signal

Exemple 2.3.2 Soit le signal analogique

$$x(t) = \sin(2\pi ft) \quad \text{avec } f = 2 \text{ Hz}$$

Donc $f_{max} = 2$. On échantillonne par exemple avec une fréquence $f_e = 8 \text{ Hz}$, donc une période $T_e = \frac{1}{f_e} = 0,125 \text{ s}$. Les échantillons $x[n]$ sont comme suit :

$$x[0] = \sin(2\pi \cdot 2 \cdot 0) = 0$$

$$x[1] = \sin(2\pi \cdot 2 \cdot 0,125) = \sin(\pi/2) = 1$$

$$x[2] = \sin(2\pi \cdot 2 \cdot 0,25) = \sin(\pi) = 0$$

$$x[3] = \sin(2\pi \cdot 2 \cdot 0,375) = \sin(3\pi/2) = -1$$

$$x[4] = \sin(2\pi \cdot 2 \cdot 0,5) = \sin(2\pi) = 0$$

etc.

Remarque 3 On pratique, on utilise un convertisseur analogique-numérique (CAN) pour transformer un signal continu $x(t)$ en une suite numérique $x_e(t) = x(nT_e)$, où T_e est la période d'échantillonnage (la figure 2.10).

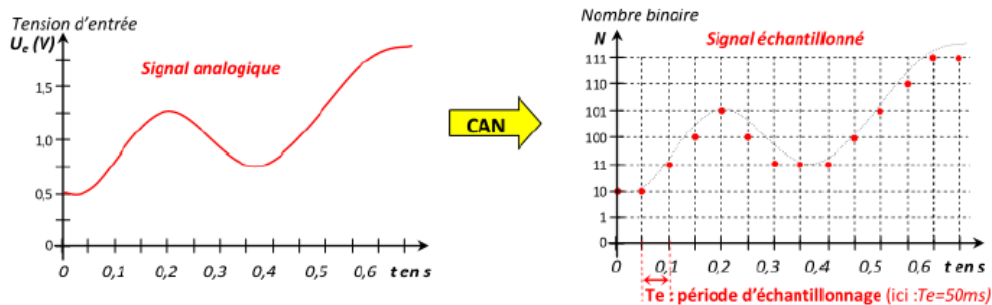


FIGURE 2.10 – Echantillonnage d'un signal analogique

(b) Quantification

La **quantification** est une étape essentielle du processus de numérisation d'un signal. Son objectif est de convertir les valeurs réelles continues du signal échantillonné $x(nT_e)$ en valeurs discrètes appartenant à un ensemble fini (la figure 2.11). Le principe consiste à arrondir chaque valeur échantillonnée au niveau le plus proche sur une échelle discrète déterminée par le pas de quantification. Cette opération simplifie la représentation numérique du signal, mais elle introduit inévitablement une *erreur de quantification*, résultant du fait que plusieurs valeurs réelles peuvent être représentées par une même valeur discrète.

(c) Codage

Le **codage** constitue l'ultime étape de la conversion d'un signal analogique en signal numérique. Il consiste à représenter chaque valeur quantifiée par un mot binaire, c'est-à-dire une suite de bits. Lorsque la quantification utilise N niveaux, le codage nécessite au minimum $\log_2(N)$ bits pour représenter chaque valeur. Par exemple, si l'on dispose de 256 niveaux de quantification, chaque échantillon est codé sur 8 bits, soit un octet.

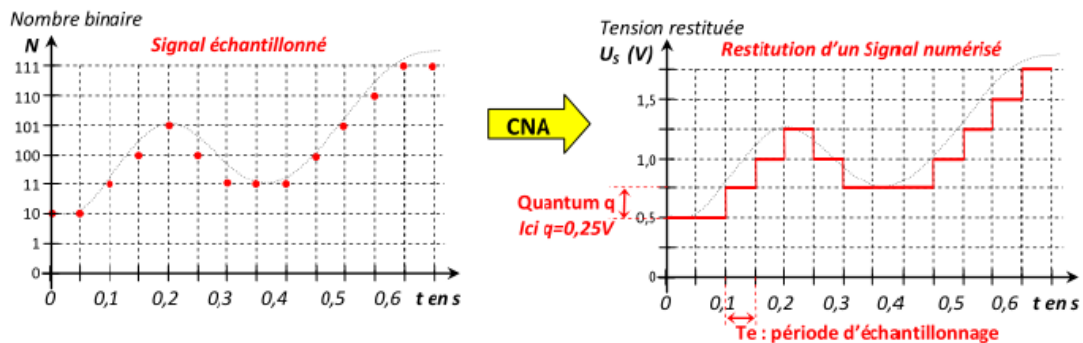


FIGURE 2.11 – Restitution d'un signal numérisé

2. Classification selon la périodicité

- **Signaux périodiques** : ce sont des signaux qui se répètent à intervalles réguliers, comme un signal sinusoïdal. En général, un signal périodique peut être exprimé mathématiquement sous la forme

$$x(t) = A \sin(\omega t + \phi),$$

où

A est l'amplitude (la valeur maximale que le signal peut atteindre) ;

ω est la fréquence angulaire (en radians par seconde), liée à la fréquence f (en Hertz) par la relation $\omega = 2\pi f$;

f est la fréquence (le nombre de cycles par seconde) ;

ϕ est la phase initiale (qui indique le décalage du signal à l'instant $t = 0$) ;

t est le temps.

Pour décomposer un signal périodique, on utilise les séries de Fourier, une méthode mathématique qui le représente comme une superposition de sinusoïdes aux fréquences, amplitudes et phases variées.

Pour un signal périodique $x(t)$ de période T , la série de Fourier est donnée par :

$$x(t) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos\left(\frac{2\pi n}{T}t\right) + b_n \sin\left(\frac{2\pi n}{T}t\right) \right),$$

où a_0 , a_n et b_n sont les coefficients de la série de Fourier, définis comme suit :

$$a_0 = \frac{1}{T} \int_T x(t) dt,$$

$$a_n = \frac{2}{T} \int_T x(t) \cos\left(\frac{2\pi n t}{T}\right) dt,$$

$$b_n = \frac{2}{T} \int_T x(t) \sin\left(\frac{2\pi n t}{T}\right) dt$$

Après avoir défini les coefficients trigonométriques de la série de Fourier, nous pouvons également considérer les coefficients exponentiels, qui offrent une représentation compacte des fonctions périodiques et facilitent l'analyse des signaux.

Pour un signal périodique $x(t)$ de période T . Les coefficients exponentiels c_n sont définis par :

$$c_n = \frac{1}{T} \int_0^x x(t) e^{-i\frac{2\pi n}{T}t} dt \quad \text{avec } n \in \mathbb{Z}$$

La série de Fourier exponentielle s'écrit alors sous la forme :

$$x(t) = \sum_{n=-\infty}^{+\infty} c_n e^{i\frac{2\pi n}{T}t}$$

où

- c_n sont les coefficients complexes,
- T est la période de la fonction,
- n est un entier relatif,
- i est l'unité imaginaire.

- **Signaux apériodiques** : ne présentent pas de répétition exacte dans le temps, comme un signal transitoire.

La transformée de Fourier permet de représenter un signal apériodique dans le domaine fréquentiel. Pour un signal $x(t)$, la transformée de Fourier est définie par :

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi f t} dt$$

où $X(f)$ est la représentation du signal dans le domaine des fréquences.

3. Classification selon la détermination

- **Signaux déterministes** : Parfaitement décrits par une expression mathématique et donc entièrement prévisibles (la figure 2.12). Par exemple, un signal sinusoïdal peut s'écrire $s(t) = A \cos(\omega t + \alpha)$. Ces signaux sont essentiels pour l'analyse des systèmes linéaires invariants dans le temps (LTI), permettant d'étudier leur réponse en fréquence $H(\omega)$ et le temps de propagation de groupe $\tau(\omega)$. Cependant, ils ne représentent pas toujours la nature aléatoire des signaux réels.
- **Signaux aléatoires** : leur amplitude $s(t)$ à chaque instant t est décrite par une loi de probabilité,

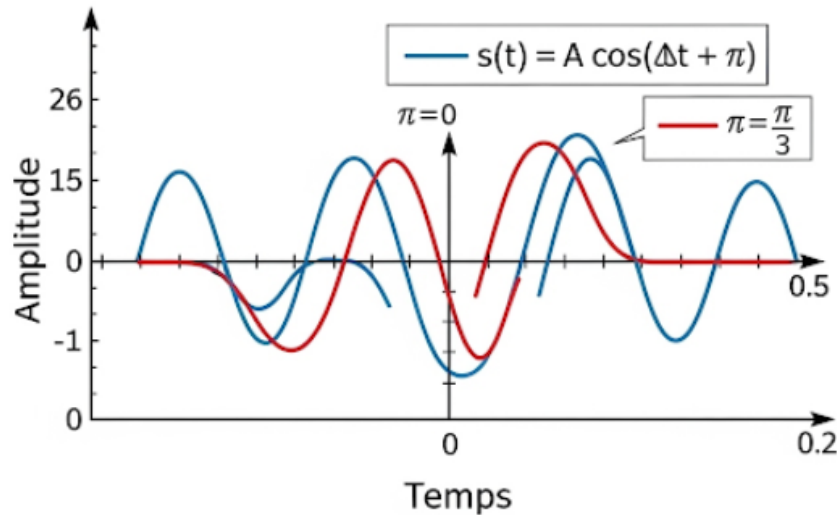


FIGURE 2.12 – Signaux déterministes

généralement exprimée par une densité de probabilité $p(x, t)$ définie par :

$$p(x, t) = \lim_{\Delta x \rightarrow 0} \frac{\text{Prob} [x \leq s(t) \leq x + \Delta x]}{\Delta x}$$

où Δx représente une petite variation ou une petite plage de valeurs pour l'amplitude du signal.

4. Selon la stationnarité :

- **Signaux stationnaires** : leurs caractéristiques statistiques (moyenne, variance, etc.) ne varient pas dans le temps, ce qui facilite leur analyse fréquentielle. Exemple, Un signal aléatoire est dit **stationnaire** si ses propriétés statistiques, notamment la densité de probabilité, restent constantes dans le temps, c'est-à-dire $p(x, t) = p(x)$.
- **Signaux non stationnaires** : leurs propriétés statistiques évoluent au cours du temps. C'est le cas de nombreux signaux réels, tels que la parole ou les signaux EEG, qui nécessitent des méthodes d'analyse spécifiques comme la transformée de Fourier à court terme (TFCT) ou la transformée en ondelettes [15].

2.4 Nature, classification et traitement du bruit

Définition et sources du bruit

Le bruit, dans le contexte du traitement du signal, désigne toute composante non désirée qui vient perturber ou masquer l'information utile contenue dans un signal. Il peut résulter de diverses sources, internes ou externes au système de mesure ou de transmission.

Parmi les sources courantes du bruit, on peut citer les problèmes inhérents aux composants électroniques (qui peuvent générer un souffle ou un bourdonnement de fond, typiques des signaux audio), les interférences électromagnétiques (causant des grésillements ou des clics audibles), ou encore les erreurs de quantification lors de la numérisation.

Le bruit est souvent modélisé comme un signal aléatoire, ce qui permet d'en étudier les propriétés statistiques et de développer des techniques pour l'éliminer ou le réduire par filtrage.

2.4.1 Impact du bruit sur la perception

Le bruit, en tant que son indésirable ou perturbateur, exerce un impact profond sur la perception auditive humaine. auditive irréversible. Ces effets se manifestent de diverses façons [11] :

- **Masquage auditif** : Le bruit ambiant peut couvrir des sons importants, rendant difficile la compréhension de la parole. Lorsque le niveau de bruit L_b dépasse celui de la parole L_p , la communication devient compromise
- **Diminution de l'intelligibilité de la parole** : Le bruit contraint les individus à élever la voix ou à répéter, ce qui altère la clarté du message et accroît la charge cognitive. Le rapport signal à bruit, qui sera défini dans ce qui suit, est ici déterminant : plus le bruit est élevé par rapport au signal utile, plus l'intelligibilité diminue.
- **Fatigue auditive** : Une exposition prolongée au bruit, même à des niveaux modérés, peut entraîner une fatigue des cellules sensorielles de l'oreille. Cela se traduit par une sensation d'oreille bouchée, des acouphènes temporaires (bourdonnements, sifflements), voire une baisse passagère de l'audition.
- **Perte auditive induite par le bruit** : Une exposition chronique ou intense à des niveaux de bruit supérieurs à 80 – 85 dB peut causer des lésions irréversibles des cellules ciliées de l'oreille interne, entraînant une surdité permanente.

2.4.2 Classification et caractéristiques du bruit

Le bruit peut être classé selon plusieurs critères : le bruit déterministe et le bruit aléatoire.

— **Bruit Déterministe**

Le bruit déterministe est un signal indésirable mais prévisible, souvent perçu comme un son. Il se distingue par la présence de fréquences précises et de leurs harmoniques, qui se répètent périodiquement. Cette régularité permet de le modéliser précisément par des équations mathématiques. Par exemple, un bourdonnement ou un ronflement provenant d'un appareil électrique (comme celui causé par le courant alternatif à 50 ou 60 Hz) est un bruit déterministe. Une composante sinusoïdale de ce bruit peut être représentée par :

$$x(t) = A \sin(2f_0 t + \phi)$$

où f_0 représente la fréquence fondamentale du signal, en Hertz (Hz).

Cette capacité à décrire le bruit par une équation permet, en théorie, de l'identifier et de le soustraire ou de le filtrer du signal utile.

— **Bruit aléatoire** Le bruit aléatoire est imprévisible; on ne peut pas connaître sa valeur exacte à un instant donné. Son amplitude à chaque instant est plutôt caractérisée par une probabilité. On le divise principalement en deux catégories : le bruit blanc et le bruit coloré.

- **Bruit blanc** : Le bruit blanc $b(t)$ est un type de bruit aléatoire idéalisé. Il se caractérise par deux propriétés clés :

- Son spectre de puissance est constant ($S_b(f) = Cte$), ce qui signifie qu'il contient toutes les fréquences en proportions égales.
- Sa fonction d'autocorrélation est un pic de Dirac ($C_{bb}(\tau) = K \cdot \delta(\tau)$), indiquant que ses valeurs à des instants différents sont totalement indépendantes.

Ce modèle est fondamental pour l'analyse théorique du bruit.

- **Bruit coloré [12]** : Le bruit coloré est un bruit aléatoire dont le spectre de puissance décroît avec la fréquence, suivant la relation :

$$S(f) = f^{-\alpha}$$

L'exposant α détermine sa "couleur". Le cas le plus connu est le bruit en $1/f$ (ou bruit rose), pour lequel $\alpha = 1$, très présent dans divers phénomènes naturels et physiques.

2.4.3 Modélisation du bruit dans un signal audio

Dans divers contextes, comme les communications audio ou les mesures scientifiques, le bruit peut nuire à la clarté et à la compréhension des signaux. L'analyse de ce phénomène à l'aide de la transformée de Fourier révèle souvent une distribution spectrale étendue et aléatoire sur différentes fréquences. Comprendre les propriétés du bruit est crucial pour développer des techniques de réduction efficaces, permettant ainsi d'améliorer la qualité des signaux et d'optimiser la transmission des informations .

Un signal audio bruité $x(t)$ peut être modélisé comme la somme de deux composantes [16] : un **signal utile** $s(t)$, contenant l'information d'intérêt (par exemple la parole), et un **bruit additif** $b(t)$, considéré comme une perturbation indésirable

$$x(t) = s(t) + b(t)$$

Après échantillonnage, ce modèle devient :

$$x[n] = s[n] + b[n]$$

- **Signal utile** $s[n]$: composante contenant les informations pertinentes du signal (voix, musique, etc.), que l'on souhaite extraire ou préserver.
- **Bruit additif** $b[n]$: signal indésirable ajouté au signal utile, dégradant sa qualité.
- **Signal bruité** $x[n]$: combinaison du signal utile et du bruit, c'est ce qui est capté par le microphone.

2.5 Bruits courants et principes de réduction associés

Le bruit peut prendre des formes variées : Stationnaire (constant dans le temps, comme le bruit d'un ventilateur), fluctuant (varie aléatoirement) ou impulsionnel (bruits soudains et courts), chaque type de bruit nécessite une stratégie de traitement adaptée pour améliorer l'intelligibilité de la parole.

Réducteur de bruit impulsionnel : Il atténue les pics soudains de pression acoustique (comme un claquement ou un choc), qui durent environ 10 millisecondes. Ce type de bruit est difficile à anticiper, d'où l'importance d'une réponse rapide du système.

Réducteur de bruit de vent : Il supprime les fluctuations de basse fréquence entre les deux oreilles, causées par le vent frappant les microphones. Ces variations sont appelées variations interaurales.

Une solution naturelle est d'exploiter la forme de l'oreille (effet pavillonnaire) : les appareils intra-auriculaires ou les écouteurs avec microphone déporté permettent de réduire cet effet sans traitement électronique complexe.

Réducteur de bruit ambiant : Ce bruit est constant et diffus (comme dans un restaurant ou une rue animée). On peut le réduire de deux façons :

- Par directivité microphonique : capter préférentiellement le son venant de face (la parole) plutôt que des côtés ou de l'arrière (le bruit).
- Par filtrage spectral : analyser le spectre du signal bruité pour y identifier la composante vocale et soustraire le spectre estimé du bruit.

L'objectif fondamental de la réduction de bruit est d'estimer $s[n]$ à partir de $x[n]$, en minimisant l'effet du bruit. Pour ce faire, plusieurs approches ont été développées.

2.5.1 Méthodes de réduction du bruit : revue des techniques

Le choix d'une technique de réduction du bruit dépend du nombre d'observations disponibles et de la nature du signal.

1. **Techniques monovoies** : cas le plus courant, où un seul canal d'observation bruitée est disponible.
2. **Techniques multivoies** : ces approches disposent d'au moins deux canaux d'observation (microphones), permettant une meilleure estimation du signal utile.
3. **Méthodes statistiques et projections spectrales** : il est possible de réduire le bruit en projetant le signal de parole bruité sur un sous-espace associé au signal utile.
4. **Techniques avancées** : Certaines approches utilisent :
 - des réseaux de neurones profonds (Deep Neural Networks),
 - des modèles de Markov cachés (HMM : Hidden Markov Models),
 - ou encore des techniques de masquage spectral.

2.5.2 Indicateurs de qualité d'un signal audio

L'évaluation de la qualité d'un signal audio traité repose sur des **indicateurs perceptifs et/ou objectifs**. Les **indicateurs perceptifs**, quant à eux, visent à modéliser la perception humaine du son, en s'appuyant sur des modèles psychoacoustiques du système auditif. Ils sont conçus pour estimer la qualité ou l'intelligibilité perçue par un auditeur humain. Parmi les plus connus figurent le *PESQ* (Perceptual Evaluation of Speech Quality) et le *STOI* (Short-Time Objective Intelligibility). Bien qu'ils soient plus proches de l'expérience subjective, leur mise en œuvre est souvent plus complexe.

Les **indicateurs objectifs** sont des mesures quantitatives calculées à partir des signaux eux-mêmes (souvent le signal original et le signal traité). Ils reposent sur des formules mathématiques précises et ne font pas appel à une appréciation humaine directe. Parmi eux, on trouve par exemple le *Rapport Signal sur Bruit (RSB)* ou le *Rapport Signal sur Distorsion (SDR)*. Ces indicateurs sont rapides à calculer et reproductibles, mais ne reflètent pas toujours la perception auditive réelle.

Rapport Signal sur Bruit

Le rapport signal sur bruit, noté RSB (ou SNR en anglais pour *Signal-to-Noise Ratio*), est une mesure classique permettant d'évaluer la qualité d'un signal audio, en comparant l'intensité du signal utile à celle du bruit présent.

Soient $s[n]$ le signal propre (sans bruit) et $b[n]$ le bruit (par exemple, le bruit résiduel après traitement), le RSB est défini par :

$$\text{RSB} = 10 \log_{10} \left(\frac{\|s[n]\|^2}{\|b[n]\|^2} \right) \quad (\text{en dB})$$

où la norme utilisée correspond à l'énergie (ou la puissance) du signal, selon le contexte. Pour un signal discret de longueur finie, cette norme est donnée par :

$$\|X[n]\|^2 = \sum_{n=0}^{N-1} |X[n]|^2$$

Pour un signal discret E , l'énergie est définie comme la somme des carrés de ses amplitudes :

$$E = \sum_{n=-\infty}^{+\infty} |X[n]|^2$$

Un signal est dit à *énergie finie* si E converge. C'est le cas de la plupart des signaux audio de durée limitée, comme les trames utilisées par la TFCT. Pour un signal de durée infinie ou stationnaire, on considère plutôt la puissance moyenne, définie par :

$$P = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N |X[n]|^2$$

Elle représente l'énergie moyenne.

Remarque 4 : dans les contextes pratiques de traitement du signal audio (fenêtrage, analyse par blocs), les mesures telles que le RSB ou le SDR utilisent généralement l'*énergie*.

Rapport Signal sur Distorsion

Le SDR est une mesure plus complète qui évalue la qualité du signal estimé après traitement en le comparant au signal propre. Contrairement au RSB, il prend en compte non seulement le bruit résiduel, mais aussi les artefacts ou distorsions introduits par le traitement.

Soient $s[n]$ le signal de référence (propre) et $\hat{s}[n]$ le signal estimé (traité), alors le SDR s'écrit :

$$\text{SDR} = 10 \log_{10} \left(\frac{\|s[n]\|^2}{\|s[n] - \hat{s}[n]\|^2} \right) \quad (\text{en dB})$$

Le dénominateur représente l'énergie de l'erreur entre le signal original et le signal reconstruit. Plus le SDR est élevé, meilleure est la fidélité du traitement.

Ces deux indicateurs sont couramment utilisés dans les évaluations de performance des algorithmes de réduction de bruit.

2.6 Conclusion

Ce chapitre a posé les bases du traitement du signal en s'appuyant sur l'analyse des signaux sonores. Il a d'abord présenté les propriétés physiques des signaux et leur classification. La représentation des signaux et l'influence du bruit sur l'audition ont ensuite été abordées. Les différentes formes de bruit et leurs caractéristiques

ont été détaillées. Une modélisation d'un signal bruité a été proposée, suivie de techniques de réduction du bruit. Enfin, les notions de RSB et RSD ont été introduites comme critères de qualité.

Chapitre 3

La Transformée de Fourier dans le traitement du signal

3.1 Introduction

La réduction du bruit est un enjeu central dans le traitement des signaux audio, en particulier pour les appareils auditifs. Elle consiste à atténuer les composantes indésirables d'un signal sans altérer le contenu utile, comme la parole. Ce processus implique un compromis délicat entre efficacité et préservation de la qualité sonore. Dans des environnements complexes, la performance des algorithmes de réduction conditionne fortement l'intelligibilité et le confort d'écoute. La transformée de Fourier joue un rôle clé dans ces méthodes en permettant une analyse fine des composantes fréquentielles du signal. Elle est à la base de nombreuses techniques classiques et modernes de débruitage. Ce chapitre explore les différentes formes de la transformée de Fourier et leur utilisation pour isoler et réduire le bruit. L'objectif est de comprendre comment la transformation fréquentielle permet une manipulation ciblée du signal audio.

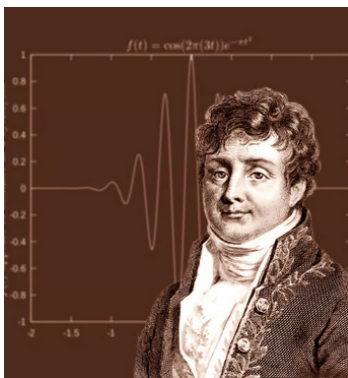


FIGURE 3.1 – Joseph Fourier

Joseph Fourier : un précurseur de l'analyse fréquentielle

Jean-Baptiste Joseph Fourier (1768–1830) fut un mathématicien et physicien français, connu pour avoir posé les bases de l'analyse fréquentielle. Après une carrière d'enseignant et de préfet, notamment à Grenoble, il se consacra à l'étude de la propagation de la chaleur dans les corps solides.

Ses travaux aboutirent à une méthode révolutionnaire : la décomposition de fonctions en séries trigonométriques, aujourd'hui connue sous le nom *d'analyse de Fourier*. Bien que critiquée à son époque, cette méthode s'est révélée fondamentale pour le traitement des signaux. Fourier fut membre influent de l'Académie des sciences. Son héritage scientifique reste incontournable (la figure 3.1).

3.2 Pourquoi le domaine fréquentiel ?

Un signal sonore est généralement composé de plusieurs fréquences superposées, comme celles de la voix, de la musique ou encore du bruit ambiant. Dans le domaine temporel, ces composantes sont entremêlées, ce qui rend leur analyse difficile. Le passage dans le domaine fréquentiel, grâce à la transformée de Fourier, permet de

décomposer le signal en une somme de sinusôides de différentes fréquences, facilitant ainsi l'identification des composantes dominantes. Par exemple, la parole humaine est principalement située entre 300 Hz et 3400 Hz, tandis que de nombreux bruits de fond se trouvent en dehors de cette bande. Une telle représentation rend alors possible la suppression ciblée des fréquences indésirables.

De plus, le bruit et le signal présentent souvent des structures spectrales distinctes. Le bruit de fond, comme celui d'un ventilateur ou d'un moteur, tend à être réparti uniformément sur l'ensemble des fréquences (spectre plat), alors que la parole présente un spectre structuré avec des pics à certaines fréquences appelées formants. Le domaine fréquentiel permet donc de distinguer efficacement le signal utile du bruit.

Par ailleurs, de nombreuses opérations deviennent plus simples et plus efficaces dans ce domaine : la convolution devient une multiplication, le filtrage s'effectue par masquage fréquentiel, et la réduction de bruit peut être réalisée en soustrayant un spectre estimé de bruit. En outre, plusieurs techniques avancées, comme la transformée de Fourier à court terme (TFCT), la soustraction spectrale, ou encore l'utilisation de réseaux de neurones convolutifs (CNN) sur des spectrogrammes, nécessitent une représentation fréquentielle préalable. Enfin, ce choix est également justifié biologiquement : l'oreille interne, via la cochlée, effectue elle-même une analyse fréquentielle naturelle du son, chaque fréquence stimulant une zone spécifique de la membrane basilaire. Ainsi, le domaine fréquentiel s'impose non seulement pour des raisons techniques, mais aussi en cohérence avec le fonctionnement physiologique de l'audition humaine.

3.3 Transformée de Fourier "TF" d'un signal continu

3.3.1 Définition et propriétés

Dans cette section, nous introduisons la définition mathématique de la transformée de Fourier d'un signal continu, puis nous présentons ses principales propriétés utiles en pratique.

La transformée de Fourier d'un signal $x(t)$ absolument intégrable est définie par

$$TF(x)(f) = X(f) = \int_{-\infty}^{+\infty} x(t) e^{-i2\pi ft} dt, \quad (3.1)$$

où f désigne la fréquence en Hz, et $X(f)$ représente le spectre fréquentiel du signal.

Notons que si $x(t)$ est un signal intégrable (c-à-d à énergie finie), alors $x(t)$ admet une TF.

Exemple 3.3.1 La transformée de Fourier du signal $x(t) = e^{-at}e(t)$, avec $a > 0$ et $e(t) = \chi_{\mathbb{R}^+}(t)$ la fonction échelon unité, est donnée par

$$X(f) = \int_0^{+\infty} e^{-at} e^{-i2\pi ft} dt = \frac{1}{a + i2\pi f}. \quad (3.2)$$

Ce résultat met en évidence la décroissance spectrale associée à un signal exponentiel décroissant.

Exemple 3.3.2 Soit le signal exponentiel symétrique est défini par :

$x(t) = e^{-a|t|}$, $a > 0$ (3.3) Ce signal est défini sur \mathbb{R} et décroît rapidement à zéro pour $t \rightarrow \pm\infty$. Sa transformée de Fourier, illustrée par la figure 3.2, est donnée par :

$$X(f) = \int_{-\infty}^{+\infty} e^{-a|t|} e^{-i2\pi ft} dt = \frac{2a}{a^2 + (2\pi f)^2} \quad (3.4)$$

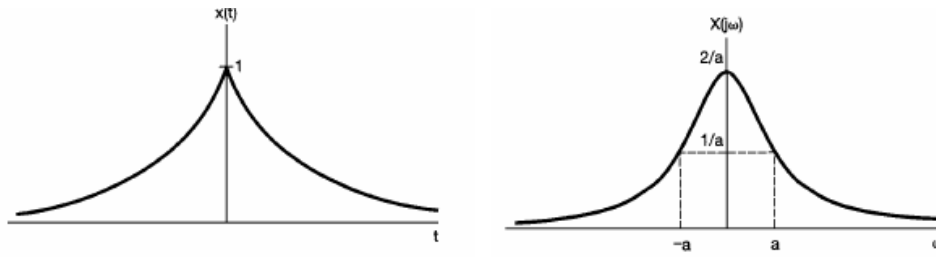


FIGURE 3.2 – Le signal $x(t) = e^{-a|t|}$ et sa transformée de Fourier

Remarque 5 La formule (3.1) de la transformée de Fourier peut également être exprimée à l'aide des fonctions trigonométriques, en utilisant la formule d'Euler

$$X(f) = \int_{-\infty}^{+\infty} x(t) [\cos(2\pi ft) - i \sin(2\pi ft)] dt. \quad (3.5)$$

On peut alors écrire la transformée comme la somme de deux composantes réelles

$$X(f) = A(f) + iB(f), \quad (3.6)$$

où

$$A(f) = \int_{-\infty}^{+\infty} x(t) \cos(2\pi ft) dt, \quad B(f) = - \int_{-\infty}^{+\infty} x(t) \sin(2\pi ft) dt. \quad (3.7)$$

Ces deux composantes permettent de définir

le spectre d'amplitude : $|X(f)| = \sqrt{A^2(f) + B^2(f)}$

le spectre de phase : $\varphi(f) = \arg(X(f)) = \arctan\left(\frac{B(f)}{A(f)}\right)$

Transformée de Fourier inverse

Sous certaines conditions d'intégrabilité et de régularité, la transformée de Fourier inverse permet de reconstruire le signal temporel $x(t)$ à partir de sa représentation fréquentielle $X(f)$. Elle s'écrit :

$$x(t) = \int_{-\infty}^{+\infty} X(f) e^{2i\pi ft} df. \quad (3.8)$$

Dans ce contexte, la transformée de Fourier et sa réciproque forment une paire d'opérations inverses, permettant une analyse et une synthèse fidèle du signal.

Propriétés algébriques de la transformée de Fourier

La transformée de Fourier (notée ici TF) possède plusieurs propriétés fondamentales qui facilitent l'analyse des signaux. Ces propriétés permettent d'interpréter les effets de certaines opérations dans le domaine temporel sur leur représentation fréquentielle.

Linéarité. La transformation de Fourier est une opération linéaire :

$$X[ax_1(t) + bx_2(t)](f) = aX_1(f) + bX_2(f)$$

où $X_1(f) = \text{TF}[x_1](f)$, $X_2(f) = \text{TF}[x_2](f)$, et $a, b \in \mathbb{R}$.

La linéarité exprime le principe de superposition, c'est-à-dire la réponse fréquentielle d'une combinaison de signaux est la combinaison des réponses fréquentielles.

Dérivation. La TF d'une dérivée temporelle est

$$\text{TF}[x'(t)](f) = 2i\pi f X(f) \quad (3.9)$$

La formule (3.9) signifie que les composantes de haute fréquence sont amplifiées (proportionnellement à f), car ce sont elles qui correspondent aux variations rapides du signal dans le domaine temporel.

En effet, la dérivée d'un signal $x(t)$ est définie par :

$$x'(t) = \lim_{\Delta t \rightarrow 0} \frac{x(t + \Delta t) - x(t)}{\Delta t},$$

elle mesure la vitesse de variation instantanée du signal à l'instant t .

Exemple 3.3.3

Pour un signal de parole la dérivée met en évidence les variations rapides du contenu vocal

- *Lorsque la parole est soutenue ou les voyelles prolongées, le signal varie lentement : la dérivée est petite.*
- *Lorsqu'un son brusque est prononcé, le signal change rapidement d'amplitude : la dérivée atteint alors des valeurs élevées.*

Parité. La symétrie d'un signal dans le domaine temporel affecte la nature de sa TF.

Si $x(t)$ est paire (i.e. $x(-t) = x(t)$), alors $X(f)$ est réelle et paire.

Si $x(t)$ est impaire (i.e. $x(-t) = -x(t)$), alors $X(f)$ est imaginaire pure et impaire.

Translation temporelle. Un décalage temporel de τ induit une modulation de phase dans le domaine fréquentiel

$$\text{TF}[x(t - \tau)](f) = e^{-i2\pi f \tau} X(f).$$

Cela signifie que traduire un signal dans le temps (le retarder ou l'avancer) ne change pas ses amplitudes spectrales $|X(f)|$, mais affecte sa phase à chaque fréquence.

Produit de convolution : Mathématiquement, le produit de convolution désigne une opération intégrale fondamentale entre deux fonctions, utilisée notamment en traitement du signal et en analyse linéaire des systèmes. Il permet de mesurer comment une fonction influence ou « filtre » l'autre au fil du temps.

Définition 2 *Étant donné g_1 et g_2 , deux fonctions intégrables au sens de Lebesgue, c'est-à-dire qu'elles appartiennent à l'espace $L^1(\mathbb{R})$ leur convolution est définie par*

$$(g_1 * g_2)(t) = \int_{-\infty}^{+\infty} g_1(\tau) g_2(t - \tau) d\tau.$$

Le produit de convolution est commutatif, $g_1 * g_2 = g_2 * g_1$; associatif $(g_1 * g_2) * g_3 = g_1 * (g_2 * g_3)$ et il est distributif par rapport à l'addition

$$g_1 * (g_2 + g_3) = g_1 * g_2 + g_1 * g_3.$$

Notons aussi que la convolution est une opération linéaire, et invariante par translation si $T_a g(t) = g(t - a)$, alors

$$T_a(g_1 * g_2) = (T_a g_1) * g_2 = g_1 * (T_a g_2)$$

Si $g_1 \in L^1(\mathbb{R})$ et $g_2 \in L^2(\mathbb{R})$, alors leur convolution $g_1 * g_2$ existe presque partout, et on a $g_1 * g_2 \in L^2(\mathbb{R})$

La convolution dans le domaine temporel devient un produit dans le domaine fréquentiel. C'est-à-dire

$$TF(g_1 * g_2)(f) = TF(g_1)(f) \cdot TF(g_2)(f) \text{ et } TF(g_1 \cdot g_2)(f) = TF(g_1 * g_2)(f)$$

Ce résultat est central en traitement du signal, car il permet d'effectuer la convolution efficacement en utilisant la transformée de Fourier rapide (TFR).

Produit de Convolution : Réponse d'un système à une entrée quelconque

Soit $x(t)$ un signal d'entrée appliqué à un système (S) linéaire et invariant dans le temps (LTI) (voir la figure 3.3), dont la réponse impulsionnelle est notée $h(t)$. La sortie correspondante à $x(t)$ est $y(t)$.

$$x(t) = \int_{-\infty}^{+\infty} x(\tau) \delta(t - \tau) d\tau \quad (3.10)$$

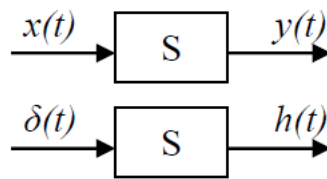


FIGURE 3.3 – Présentation d'un système LTI

Le signal $x(t)$ peut être représenté comme une somme infinie de contributions pondérées d'impulsions de Dirac

Posons g l'application linéaire qui, à toute entrée $x(t)$, associe la sortie $y(t)$ du système

$$g : x(t) \mapsto y(t) = g(x(t)) \quad (3.11)$$

Ainsi, on a :

$$y(t) = g(x(t)) = g\left(\int_{-\infty}^{+\infty} x(\tau) \delta(t - \tau) d\tau\right). \quad (3.12)$$

Par linéarité de g , l'expression (3.11) devient :

$$y(t) = \int_{-\infty}^{+\infty} x(\tau) g(\delta(t - \tau)) d\tau. \quad (3.13)$$

Par définition, la réponse impulsionnelle $h(t - \tau)$ est la sortie du système à une impulsion retardée $\delta(t - \tau)$

$$g(\delta(t - \tau)) = h(t - \tau). \quad (3.14)$$

D'où l'expression finale de la sortie $y(t)$ comme un produit de convolution

$$y(t) = \int_{-\infty}^{+\infty} x(\tau) h(t - \tau) d\tau = x(t) * h(t). \quad (3.15)$$

Remarque 6 La TF fournit ainsi une description complète d'un signal en termes de ses composantes fréquentielles. Toutefois, elle présente certaines limites : elle suppose que le signal est défini sur tout l'intervalle $]-\infty, +\infty[$, ce qui est irréaliste pour des signaux réels souvent bornés, et elle ne fournit aucune information temporelle sur l'apparition des fréquences.

Cette absence de localité ne constitue pas un inconvénient majeur pour les signaux stationnaires, c'est-à-dire ceux dont la structure ne varie pas significativement dans le temps. En revanche, pour les signaux non stationnaires, dont le contenu fréquentiel évolue, cette perte d'information temporelle devient problématique. Cette transformation masque alors les variations temporelles du signal, empêchant une analyse fine de son évolution dynamique. C'est dans ce contexte que des approches alternatives, comme la transformée de Fourier à fenêtre (TFCT) ou la transformée en ondelettes, ont été développées afin de mieux capter simultanément les informations en temps et en fréquence.

3.4 Transformée de Fourier Discrète

Du point de vue mathématique la transformée de Fourier discrète est un simple changement de base dans un espace vectoriel, néanmoins, son interprétation a une importance et une utilité énorme dans les applications, en particulier en théorie de signal. L'étude de la (TFD) a fortement contribué au développement du monde numérique notamment aux domaines l'audiovisuel et l'imagerie médicale.

La transformée de Fourier discrète est une méthode qui permet de d'écrire un signal discret en termes de fréquence, tout comme la TF permet de d'écrire un signal continu en termes de fréquence.

3.4.1 Définitions et exemples

La TFD permet de représenter un signal discret $x[n]$ de taille N dans une base de fonctions complexes exponentielles.

Il existe deux formes équivalentes de la TFD, qui diffèrent par un facteur de normalisation.

Définition 3 *Considérons une suite numérique finie $\{x(0), x(1), \dots, x(N-1)\}$. Sa TFD est définie comme une autre suite finie, $\{X(0), X(1), \dots, X(N-1)\}$, donnée par les formules suivantes :*

Forme sans normalisation (forme algorithmique)

$$TFD(x)[k] = X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-i\frac{2\pi}{N}kn}, \quad n, k \in \{0, \dots, N-1\}. \quad (3.16)$$

Cette forme est couramment utilisée dans les implémentations pratiques (par exemple en Python ou MATLAB), car elle simplifie le calcul. La normalisation est ensuite appliquée dans la transformée inverse.

Forme avec normalisation (forme théorique)

$$X[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \cdot e^{-i\frac{2\pi}{N}kn}, \quad n, k \in \{0, \dots, N-1\}.$$

Exemple 3.4.1 *Soit $x[n] = (2, 3, 4, 4)$, la $TFD(x)$ est définie par*

$$X(k) = \sum_{n=0}^{N-1} x[n]e^{-i\frac{2\pi}{N}nk} \quad \text{pour } k = 0, 1, \dots, N-1$$

où N est le nombre d'échantillons. Dans cet exemple on prend, $N = 4$.

Calculons $X[k]$ pour $k = 0, 1, 2, 3$

$$X[0] = \sum_{n=0}^3 x[n]e^0 = 2 + 3 + 4 + 4 = 13$$

$$X[1] = 2e^0 + 3e^{-i\frac{\pi}{2}} + 4e^{-i\pi} + 4e^{-i\frac{3\pi}{2}} = 2 - 3i - 4 + 4i = -2 + i$$

$$X[2] = 2e^0 + 3e^{-i\pi} + 4e^{-i2\pi} + 4e^{-i3\pi} = 2 - 3 + 4 - 4 = -1$$

$$X[3] = 2e^0 + 3e^{-i\frac{3\pi}{2}} + 4e^{-i3\pi} + 4e^{-i\frac{9\pi}{2}} = 2 + 3i - 4 - 4i = -2 - i$$

Finalement,

$$TFDx[n] = X[k] = (13, -2 + i, -1, -2 - i).$$

3.4.2 La transformée de Fourier discrète inverse

La transformée de Fourier discrète inverse (TFDi) permet de reconstituer un signal temporel discret $x[n]$ à partir de sa représentation fréquentielle $X[k]$. Pour un signal de longueur N , la formule est donnée par

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{i\frac{2\pi}{N}kn}, \quad \text{où } n = 0, 1, \dots, N-1. \quad (3.17)$$

La TFD et la TFDi de la figure 3.4 sont parfaitement réciproques : La TFD permet de passer du domaine temporel au domaine fréquentiel, et la TFDi permet de revenir au signal d'origine sans perte d'information.

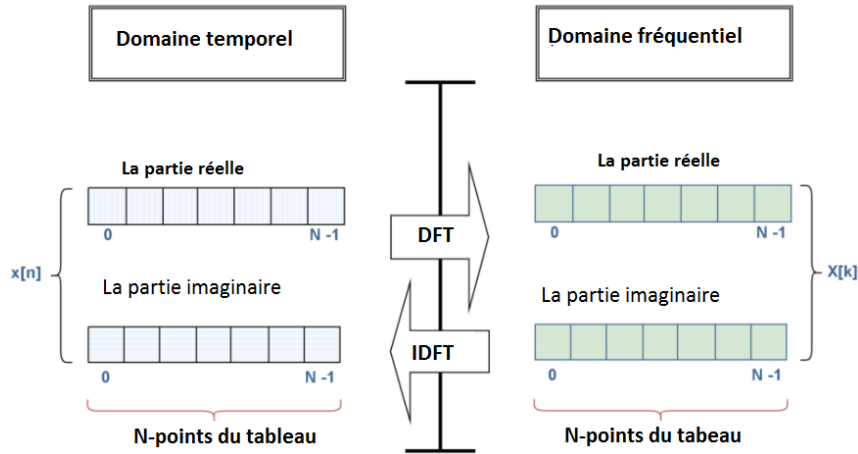


FIGURE 3.4 – TFD et TFDi

Exemple 3.4.2 Soit la séquence fréquentielle suivante pour $N = 4$:

$$X(k) = (13, -2 + i, -1, -2 - i)$$

Calculons les $x[n]$, en utilisant

Pour $n = 0$:

$$x[0] = \frac{1}{4} (13 + (-2 + i) + (-1) + (-2 - i)) = \frac{1}{4} \cdot 8 = 2 .$$

Pour $n = 1$:

$$\begin{aligned} x[1] &= \frac{1}{4} \left(13 + (-2 + i)e^{i\frac{2\pi}{4} \cdot 1} + (-1)e^{i\frac{2\pi}{4} \cdot 2} + (-2 - i)e^{i\frac{2\pi}{4} \cdot 3} \right) \\ &= \frac{1}{4} (13 + (-2 + i)i + (-1)(-1) + (-2 - i)(-i)) = 3 . \end{aligned}$$

Pour $n = 2$

$$\begin{aligned} x[2] &= \frac{1}{4} (13 + (-2 + i)(-1) + (-1)(1) + (-2 - i)(-1)) \\ &= \frac{1}{4} (13 + (2 - i) - 1 + (2 + i)) = \frac{1}{4}(16) = 4 . \end{aligned}$$

Pour $n = 3$

$$\begin{aligned} x[3] &= \frac{1}{4} (13 + (-2 + i)(-i) + (-1)(-1) + (-2 - i)(i)) \\ &= \frac{1}{4} (13 + (2i + 1) + 1 + (-2i + 1)) = \frac{1}{4}(16) = 4 . \end{aligned}$$

Ainsi la séquence temporelle reconstituée est

$$x[n] = (2, 3, 4, 4)$$

3.4.3 Propriétés de la TFD

La TFD possède les propriétés suivantes :

- **Linéarité :**

Pour deux signaux $x_1[n]$ et $x_2[n]$ de TFD respectives $X_1[k]$ et $X_2[k]$, et pour tout $a, b \in \mathbb{C}$

$$ax_1[n] + bx_2[n] \xrightarrow{\text{TFD}} aX_1[k] + bX_2[k]$$

- **Périodicité**

Le spectre TFD est périodique de période N (longueur du signal)

$$X[k + N] = X[k], \quad \forall k \in \mathbb{Z}$$

Convolution : Cas Discret

Soit $x(t)$ et $h(t)$ deux signaux discrets $x[n]$ et $h[n]$ La convolution discrète s'écrit

$$y[n] = \sum_{m=-\infty}^{+\infty} x[m] h[n - m] . \tag{3.18}$$

Dans les systèmes physiques réels, les signaux sont souvent nuls pour les indices négatifs. On a donc

$$y[n] = \sum_{m=0}^n x[m] h[n - m] \tag{3.19}$$

TFD de la convolution :

La convolution dans le domaine temporel devient un produit dans le domaine fréquentiel

$$y[n] = x[n] * h[n] \xrightarrow{\text{TFD}} Y[k] = X[k] \cdot H[k].$$

Pour que cette propriété soit valide, il est important que les signaux soient correctement étendus par zéro ou par chevauchement.

3.4.4 Interprétation matricielle de la TFD et de la TFDi

La linéarité de la TFD permet d'exprimer cette dernière sous une forme matricielle, offrant une perspective enrichissante pour son analyse et son implémentation.

La TFD d'un signal temporel discret $x[n]$ peut être exprimée sous forme d'une simple multiplication matricielle

$$X = W_N \cdot x$$

où W_N est la matrice de la TFD de taille $N \times N$, dont chaque élément à la ligne k et la colonne n est donné par $\omega_{k,n} = e^{-i\frac{2\pi}{N}kn}$.

Si on note $\omega_N = e^{-\frac{2i\pi}{N}}$, la racine N -ième de l'unité, la formule de la TFD peut s'écrire comme

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot \omega_N^{kn}, \quad k = 0, 1, \dots, N-1,$$

et la matrice de la TFD $(W_N)_{k,n} = \omega_N^{kn}$ s'écrit explicitement

$$W_N = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega_N & \omega_N^2 & \dots & \omega_N^{N-1} \\ 1 & \omega_N^2 & \omega_N^4 & \dots & \omega_N^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega_N^{N-1} & \omega_N^{2(N-1)} & \dots & \omega_N^{(N-1)(N-1)} \end{pmatrix}.$$

W_N est une matrice de Vandermonde.

En utilisant $\omega_N^{-1} = e^{\frac{2i\pi}{N}}$ (le conjugué de ω_N), on montre que la matrice de la TFDi est donnée par l'inverse de W_N , $W_N^{-1} = \frac{1}{N}W_N^*$, où W_N^* est l'adjointe (transposée de la comatrice) de W_N . Les éléments de W_N^{-1} sont $(\frac{1}{N})\omega_N^{-kn}$.

$$W_N^{-1} = \frac{1}{N} \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega_N^{-1} & \omega_N^{-2} & \dots & \omega_N^{-(N-1)} \\ 1 & \omega_N^{-2} & \omega_N^{-4} & \dots & \omega_N^{-2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega_N^{-(N-1)} & \omega_N^{-2(N-1)} & \dots & \omega_N^{-(N-1)(N-1)} \end{pmatrix}.$$

La reconstruction du signal temporel s'effectue alors par :

$$\mathbf{x} = W_N^{-1} \mathbf{X} = \frac{1}{N} W_N^* \mathbf{X} .$$

Cette représentation matricielle fournit un cadre conceptuel puissant pour l'analyse de la TFD, facilitant la compréhension de ses propriétés fondamentales et servant de base aux algorithmes de calcul efficaces comme la TFR.

Exemple 3.4.3 Pour $N = 4$: $e^{-2i\pi/4} = e^{-2i\pi/4} = -i$, donc

$$W_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & (-i)^2 & (-i)^3 \\ 1 & (-i)^2 & (-i)^4 & (-i)^6 \\ 1 & (-i)^3 & (-i)^6 & (-i)^9 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{pmatrix} .$$

Par conséquent, la matrice inverse est

$$W_4^{-1} = \frac{1}{4} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix} .$$

Exemple 3.4.4 Considérons la séquence temporelle suivante, de taille $N = 4$:

$$x = (x(0), x(1), x(2), x(3))^t = (2, 3, 4, 4)^t$$

La TFD peut s'écrire sous forme matricielle suivante

$$X = W_4 \cdot x \quad .$$

où W_4 est la matrice définie par

$$(W_4)_{k,n} = w^{kn} \quad \text{avec} \quad w = e^{-i\frac{2\pi}{4}} = e^{-i\frac{\pi}{2}} = -i \quad .$$

On obtient donc la matrice

$$W_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{pmatrix} .$$

Le produit matriciel s'écrit

$$X = W_4 \cdot x = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 3 \\ 4 \\ 4 \end{pmatrix} = \begin{pmatrix} 13 \\ -2 + i \\ -1 \\ -2 - i \end{pmatrix}.$$

3.4.5 Transformée de Fourier rapide

La transformée de Fourier rapide (TFR) en anglais Fast Fourier Transform (FFT) exploite la symétrie et la périodicité des fonctions exponentielles complexes pour réduire le nombre de calculs nécessaires. Plutôt que de calculer directement la TFD par définition, la TFR divise le problème en sous-problèmes plus petits, ce qui permet un calcul récursif.

Rappelons que pour un signal $x[n]$ de longueur N , la TFD s'écrit

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-i \frac{2\pi}{N} kn}.$$

En utilisant la TFR, cette somme se décompose en termes de la TFD des sous-signaux formés des échantillons pairs et impairs

$$X[k] = X_{\text{pairs}}[k] + e^{-i \frac{2\pi}{N} k} X_{\text{impairs}}[k]$$

où $X_{\text{pairs}}[k]$ est la TFD des échantillons d'indice pair, et $X_{\text{impairs}}[k]$ celle des échantillons d'indice impair.

Algorithme de la TFR

L'algorithme de la TFR dû à Cooley-Tukey[7] est un algorithme utilisé pour calculer efficacement la TFD d'un signal.

Nous introduisons ici les notations nécessaires pour présenter les étapes de l'algorithme TFR.

N : longueur du signal, qui doit être une puissance de 2 pour que l'algorithme de Cooley-Tukey soit applicable.

$x[n]$: signal d'entrée discret, avec $n = 0, 1, \dots, N - 1$.

$X[k]$: résultat de la TFD, avec $k = 0, 1, \dots, N - 1$.

Étapes de l'algorithme

— **Initialisation** : vérifier que N est une puissance de 2. Si ce n'est pas le cas, compléter le signal avec des zéros jusqu'à la prochaine puissance de 2.

— **Division** : Séparer le signal d'entrée $x[n]$ en deux sous-signaux constitués respectivement des indices pairs et impairs :
$$\begin{cases} x_{\text{pairs}}[k] = x[2k], & k = 0, \dots, \frac{N}{2} - 1 \\ x_{\text{impairs}}[k] = x[2k + 1], & k = 0, \dots, \frac{N}{2} - 1 \end{cases}$$

— **Récursion** : appliquer la TFD récursivement sur ces deux sous-signaux : $X_{\text{pairs}} = \text{TFR}(x_{\text{pairs}})$, $X_{\text{impairs}} = \text{TFR}(x_{\text{impairs}})$

— **Combinaison** : Combiner les résultats pour reconstituer la TFD complète de $x[n]$:

$$\text{Pour } k = 0, 1, \dots, \frac{N}{2} - 1, \begin{cases} X[k] = X_{\text{pairs}}[k] + e^{-i\frac{2\pi}{N}k} \cdot X_{\text{impairs}}[k] \\ X[k + \frac{N}{2}] = X_{\text{pairs}}[k] - e^{-i\frac{2\pi}{N}k} \cdot X_{\text{impairs}}[k] \end{cases}$$

— **Résultat** : Le tableau $X[k]$ ainsi calculé contient les coefficients de la TFD du signal d'entrée $x[n]$.

Notons que l'algorithme TFR possède une complexité en temps de calcul de l'ordre de $O(N \log N)$ ce qui est bien plus efficace que la méthode TFD d'ordre $O(N^2)$.

3.5 Transformée de Fourier à court terme : Approche fréquentielle et segmentation temporelle

La transformée de Fourier à court terme (TFCT), également appelée transformée de Fourier locale ou transformée de Fourier à fenêtre glissante, (en anglais Short-Time Fourier Transform, STFT), est une méthode d'analyse temps-fréquence utilisée pour examiner l'évolution spectrale d'un signal non stationnaire, c'est-à-dire un signal dont le contenu fréquentiel varie au cours du temps, comme la parole, la musique ou certains signaux biomédicaux [1].

Elle est particulièrement adaptée à l'étude des signaux quasi-stationnaires, comme la parole, dont les caractéristiques évoluent lentement dans le temps. Elle repose sur un découpage du signal en trames temporelles successives, afin d'en analyser localement le contenu fréquentiel. L'objectif de cette section est de présenter les différentes étapes de la mise en œuvre de la TFCT pour les signaux discrets.

3.5.1 Principes de la transformée de Fourier à court terme

Pour comprendre le fonctionnement de la TFCT, il est essentiel d'en exposer les principes fondamentaux. Cette méthode repose sur une combinaison d'analyse temporelle et fréquentielle, réalisée à travers les étapes suivantes :

- **Fenêtrage** : Le signal $x[n]$ est divisé en courts segments (ou « fenêtres »), qui peuvent se chevaucher ou non. Chaque segment est multiplié par une fonction de fenêtre (exemple fenêtre de Hamming), ce qui permet de localiser l'analyse dans le temps et de limiter les effets de bord lors de la transformation de Fourier .
Chaque segment est multiplié par une fonction de fenêtre, afin de localiser l'analyse dans le temps et de réduire les effets de bord lors de la transformation de Fourier.
- **Transformation de Fourier locale** : La TFD est appliquée à chaque segment fenêtré, produisant un spectre de fréquences pour chaque intervalle de temps.
- **Représentation temps-fréquence** : En déplaçant la fenêtre le long du signal et en répétant l'analyse, on obtient une représentation bidimensionnelle montrant l'évolution des composantes fréquentielles au fil du temps.

Nous allons à présent approfondir cette première étape clé : Le fenêtrage, dont le choix a une influence directe sur la précision temporelle et fréquentielle de l'analyse.

3.5.2 Fenêtrage et analyse spectrale

L'analyse fréquentielle d'un signal à l'aide de la TFCT repose sur l'application d'une fenêtre temporelle glissante. Cette opération consiste à multiplier le signal par une fonction appelée *fenêtre*, définie sur un intervalle de durée finie. Cela permet d'isoler localement une portion du signal et d'en extraire le contenu fréquentiel via une TF appliquée à ce segment.

Le choix de la fenêtre est une étape cruciale, car il affecte directement la précision temporelle et fréquentielle de l'analyse. Une fenêtre trop large améliore la résolution fréquentielle mais dégrade la localisation temporelle ; à l'inverse, une fenêtre courte donne une meilleure localisation temporelle mais une moins bonne résolution en fréquence.

Ce compromis est particulièrement important dans le cadre de méthodes telles que la soustraction spectrale qui sera exposée et appliquée au chapitre quatre, où une bonne séparation entre signal utile et bruit est essentielle. Une fenêtre mal choisie peut introduire des artefacts ou fausser l'estimation spectrale, compromettant ainsi l'efficacité du traitement.

Dans ce qui suit, nous présentons les fenêtres classiques couramment utilisées en analyse spectrale.

3.5.2.1 Fenêtre rectangulaire

La fenêtre rectangulaire offre une excellente résolution fréquentielle pour les signaux dont les composantes ont des amplitudes similaires. Cependant, dans le contexte des aides auditives, elle s'avère peu adaptée car elle ne permet pas de bien isoler les composantes faibles de la parole lorsqu'elles sont masquées par des bruits plus intenses. Cela peut nuire à la détection précise des sons utiles dans des environnements acoustiques complexes. La fenêtre rectangulaire est définie par

$$w[n] = \begin{cases} 1 & \text{si } 0 \leq n \leq N - 1 \\ 0 & \text{sinon} \end{cases}$$

où N : nombre total d'échantillons utilisés dans la fenêtre d'analyse.

Elle conserve l'intégralité du signal dans l'intervalle $[0, N - 1]$. Bien qu'elle offre une bonne résolution fréquentielle, elle introduit une fuite spectrale importante, ce qui peut dégrader la qualité de l'analyse dans les systèmes d'aides auditives.

3.5.2.2 Fenêtres de Hamming et Hanning

Les fenêtres de Hamming et de Hanning sont couramment utilisées dans le traitement des signaux à bande étroite, comme ceux de la parole. Leur principale qualité réside dans leur capacité à atténuer progressivement le signal aux extrémités, ce qui réduit considérablement la fuite spectrale. Cette propriété est particulièrement précieuse dans le contexte des aides auditives, où une analyse fine du contenu fréquentiel est essentielle pour améliorer l'intelligibilité de la parole en milieu bruyant. En permettant une meilleure séparation des composantes fréquentielles, ces fenêtres facilitent la distinction entre la parole et le bruit ambiant, même lorsque les conditions acoustiques sont complexes.

Fenêtre de Hamming (1915–1998)¹ La fenêtre de Hamming est définie par la formule suivante :

$$w[n] = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \leq n \leq N-1$$

Elle est couramment utilisée dans le traitement des signaux pour réduire la fuite spectrale tout en conservant une bonne résolution fréquentielle. Cela la rend utile dans les applications telles que les aides auditives, où l'on cherche à préserver l'intelligibilité de la parole dans un environnement bruyant.

Fenêtre de Hann (ou Hanning) (1839–1921)² La fenêtre de Hann est donnée par :

$$w[n] = 0.5 \left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right), \quad 0 \leq n \leq N-1$$

3.5.3 Fenêtrage temporel et définition de la TFCT

La (TFCT) consiste à découper un signal en courtes portions temporelles appelées *trames* (ou fenêtres). Sur chaque trame, on applique une (TFD), ce qui permet d'estimer le contenu fréquentiel localisé dans le temps. Ainsi, la TFCT fournit une représentation conjointe temps–fréquence du signal, révélant ses composantes spectrales (amplitudes et phases) dans des intervalles temporels successifs. Contrairement à la transformée de Fourier classique, qui donne une vision globale et stationnaire du spectre, la TFCT permet de suivre les variations locales du contenu fréquentiel.

Remarque 7

Comme pour la TF, il existe deux formulations principales de la TFCT (continue et discrète)

- **Versión continue** : elle est définie pour des signaux $x(t) \in L^2(\mathbb{R})$. La TFCT est alors donnée par

$$W_x(f, t_0) = \int_{-\infty}^{+\infty} x(t) \overline{\omega(t-t_0)} e^{-2i\pi ft} dt$$

où $\omega(t)$ est une fonction fenêtre centrée (souvent normalisée), f la fréquence, et t_0 l'instant d'analyse. Cette version est essentiellement utilisée dans les cadres théoriques.

- **Versión discrète** : elle repose sur l'échantillonnage du signal temporel et sur l'application de la TFD à des trames successives. Plus précisément, un signal discret $x[n]$ est découpé en trames de longueur M , chacune multipliée par une fenêtre $w[n]$, puis analysée par TFD. Le résultat est une représentation bidimensionnelle du signal dans le plan temps–fréquence.

*Dans ce mémoire, nous nous concentrons sur la **versión discrète** de la TFCT, qui est adaptée au traitement numérique des signaux échantillonnés (parole, signal audio, etc.).*

La section qui suit développe les principes mathématiques et les notations utilisées pour la TFCT discrète.

3.5.4 Définition mathématique de la TFCT discrète

Soit $x[n]$ un signal discret de longueur finie N , représentant une suite d'échantillons temporels. On souhaite analyser localement son contenu fréquentiel en découpant le signal en trames temporelles successives à l'aide d'une fenêtre de longueur $M \ll N$ (M est beaucoup plus petit que N).

1. Richard Hamming était un mathématicien et ingénieur américain ayant travaillé aux laboratoires Bell. Il est surtout connu pour le code correcteur d'erreurs qui porte son nom, ainsi que pour plusieurs contributions fondamentales en traitement du signal, dont la fenêtre de Hamming. Son travail a eu un impact majeur sur l'informatique, les télécommunications et l'ingénierie.

2. Julius von Hann était un météorologue autrichien. Bien qu'il n'ait pas directement conçu cette fenêtre, son nom a été donné à la fenêtre de Hann en son honneur. L'appellation «Hanning» est une forme altérée, probablement issue d'une combinaison entre «Hann» et «windowing». Le nom correct reste «fenêtre de Hann», même si l'usage courant en informatique emploie «Hanning».

Chaque trame est extraite en faisant glisser une fenêtre $w[n]$ de taille M le long du signal. Si m désigne le numéro de la trame et R le décalage (hop size) entre deux trames successives, la TFCT est définie par

$$X[m, k] = \sum_{n=0}^{M-1} x[mR + n] \cdot w[n] \cdot e^{-i\frac{2\pi}{M}kn}, \quad k \in \{0, 1, \dots, M-1\} \quad (3.20)$$

où m est l'indice de trame ($m = 0, 1, 2, \dots$);

R est le décalage (hop size) entre deux trames successives. Il s'agit de la distance, en nombre d'échantillons, entre le début de deux fenêtres successives. Une valeur faible de R implique un fort recouvrement entre trames, tandis qu'une valeur plus grande donne une analyse plus rapide mais plus grossière;

k est l'indice fréquentiel de la TFD locale, $k \in \{0, \dots, M-1\}$;

$X[m, k]$ est le coefficient de la TFCT de la trame m à la fréquence discrète k .

À chaque trame m , on extrait le segment $x[mR + n]$, on le multiplie par la fenêtre $w[n]$, puis on applique une TFD sur cette trame pondérée.

3.5.5 Écriture matricielle de la TFCT

L'ensemble des coefficients $X[m, k]$ obtenus en appliquant la TFCT à chaque trame du signal peut être organisé sous forme d'une matrice \mathbf{X} complexes, où P représente le nombre total de trames extraites du signal et M la longueur de chaque trame (dite aussi la taille de la TFD appliquée).

La matrice résultante est donnée par

$$\mathbf{X} = \begin{pmatrix} X[0, 0] & X[0, 1] & \dots & X[0, M-1] \\ X[1, 0] & X[1, 1] & \dots & X[1, M-1] \\ \vdots & \vdots & \ddots & \vdots \\ X[P-1, 0] & X[P-1, 1] & \dots & X[P-1, M-1] \end{pmatrix}$$

Chaque ligne correspond à la TFD d'une trame (contenu fréquentiel local à un instant donné), et chaque colonne donne l'évolution dans le temps d'une fréquence discrète donnée.

Remarque 8

Si le signal d'origine est de longueur N , que la fenêtre est de taille M , et que le saut entre trames (hop size) est R , alors le nombre total de trames est donné par

$$P = \left\lfloor \frac{N - M}{R} \right\rfloor + 1$$

Cette formule permet de déterminer combien de fois on peut glisser une fenêtre de taille M sur le signal, avec un décalage de R échantillons entre chaque position.

Prenons par exemple un signal de longueur $N = 1000$, une fenêtre de taille $M = 256$ et un saut $R = 128$, on obtient

$$P = \left\lfloor \frac{1000 - 256}{128} \right\rfloor + 1 = 6 \text{ trames}$$

Exemple numérique d'application de la TFCT

Exemple 3.5.1 *Considérons un signal discret de longueur 8 :*

$$x[n] = (1, 2, 0, -1, 1, 0, 2, -1) = (x[0], x[1], x[2], x[3], x[4], x[5], x[6], x[7]).$$

On choisit une fenêtre de Hamming de taille $M = 4$, définie par :

$$w[n] = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{M-1}\right), & \text{si } 0 \leq n < M \\ 0, & \text{sinon} \end{cases} \quad (3.21)$$

Les résultats sont approximés par

$$w[n] = [0.54, 0.85, 0.85, 0.54]$$

Le pas de trame R est fixé à $R = 2$, ce qui signifie que la fenêtre d'analyse est déplacée de 2 échantillons à chaque itération. Autrement dit, chaque trame commence 2 indices plus loin que la précédente. On peut ainsi extraire deux trames de taille 4 : $m = 0$ et $m = 1$.

Étant donné que le signal $x[n]$ contient $N = 8$ échantillons, et que la fenêtre $w[n]$ est de longueur $M = 4$, le nombre total de trames que l'on peut extraire est donné par :

$$P = \left\lfloor \frac{N - M}{R} \right\rfloor + 1 = \left\lfloor \frac{8 - 4}{2} \right\rfloor + 1 = 2$$

Nous obtenons donc deux trames : Trame $m = 0$: échantillons $x[0], x[1], x[2], x[3]$ et trame $m = 1$: échantillons $x[2], x[3], x[4], x[5]$

Chaque trame est ensuite multipliée élément par élément par la fenêtre $w[n]$, puis soumise à une TFD de taille $M = 4$.

Trame 1 : $m = 0$

$$x_0[n] = x[0 + n] = [1, 2, 0, -1]$$

Fenêtrage :

$$x_0[n] \cdot w[n] = [1 \cdot 0.54, 2 \cdot 0.85, 0 \cdot 0.85, -1 \cdot 0.54] = [0.54, 1.70, 0, -0.54]$$

On applique la TFD (taille 4) :

$$X[0, k] = \sum_{n=0}^3 x_0[n] \cdot w[n] \cdot e^{-i\frac{2\pi}{4}kn}, \quad k = 0, 1, 2, 3$$

- $X[0, 0] = 0.54 + 1.70 + 0 - 0.54 = 1.70$
- $X[0, 1] \approx 0.54 + 1.70(-i) + 0 + 0.54i = 0.54 - 1.16i$
- $X[0, 2] = 0.54 - 1.70 + 0 + 0.54 = -0.62$
- $X[0, 3] \approx 0.54 + 1.70(i) + 0 - 0.54i = 0.54 + 1.16i$

Trame 2 : $m = 1$

$$x_1[n] = x[2 + n] = [0, -1, 1, 0]$$

Fenêtrage :

$$x_1[n] \cdot w[n] = [0 \cdot 0.54, -1 \cdot 0.85, 1 \cdot 0.85, 0 \cdot 0.54] = [0, -0.85, 0.85, 0]$$

TFD :

- $X[1, 0] = 0 - 0.85 + 0.85 + 0 = 0$
- $X[1, 1] = 0 - 0.85(-i) + 0.85(-1) + 0 = 0.85i - 0.85$
- $X[1, 2] = 0 + 0.85 - 0.85 + 0 = 0$
- $X[1, 3] = 0 - 0.85(i) + 0.85(-1) + 0 = -0.85i - 0.85$

Écriture matricielle de la TFD

Pour chaque trame, on applique la TFD en utilisant la matrice W_4 , définie par

$$(W_4)_{k,n} = e^{-i\frac{2\pi}{4}kn}, \quad \text{pour } k, n \in \{0, 1, 2, 3\}$$

définie explicitement comme suit :

$$W_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{pmatrix}$$

Chaque trame pondérée $x_m[n] \cdot w[n]$ est donc vue comme un vecteur colonne, et sa TFD est obtenue par multiplication matricielle :

$$X[m, :] = W_4 \cdot (x_m[n] \cdot w[n]).$$

Ce qui donne,

(i) Pour $m = 0$ (trame 1)

- $X[0, 0] = 0.54 + 1.70 + 0 - 0.54 = 1.70$
- $X[0, 1] \approx 0.54 - 1.70i + 0 + 0.54i = 0.54 - 1.16i$
- $X[0, 2] = 0.54 - 1.70 + 0 + 0.54 = -0.62$
- $X[0, 3] \approx 0.54 + 1.70i + 0 - 0.54i = 0.54 + 1.16i$

(ii) Pour $m = 1$ (trame 2)

- $X[1, 0] = 0 - 0.85 + 0.85 + 0 = 0$
- $X[1, 1] = 0 + 0.85i - 0.85 + 0 = -0.85 + 0.85i$
- $X[1, 2] = 0 + 0.85 - 0.85 + 0 = 0$
- $X[1, 3] = 0 - 0.85i - 0.85 + 0 = -0.85 - 0.85i$

Finalement, la matrice de la TFCT (approximative)

$$\mathbf{X} = \begin{pmatrix} 1.70 & 0.54 - 1.16i & -0.62 & 0.54 + 1.16i \\ 0 & -0.85 + 0.85i & 0 & -0.85 - 0.85i \end{pmatrix}$$

Cette matrice \mathbf{X} représente la TFCT du signal initial. Chaque ligne correspond à une trame temporelle, et chaque colonne suit l'évolution d'une composante fréquentielle donnée dans le temps.

3.6 Conclusion

Les stratégies de traitement du signal, notamment le filtrage spectral et l'analyse par la TF, sont essentielles pour le développement et l'optimisation des aides auditives. Elles permettent non seulement d'améliorer la clarté du son, mais aussi de créer une expérience auditive personnalisée qui répond aux besoins spécifiques de chaque utilisateur. Ces techniques avancées jouent un rôle clé dans l'avenir des dispositifs auditifs, en intégrant la technologie et la science pour améliorer la qualité de vie des personnes malentendantes.

Chapitre 4

Réduction de Bruit par la soustraction spectrale : Classique et intelligente

4.1 Introduction

Ce chapitre explore la réduction de bruit par soustraction spectrale, depuis son principe de base jusqu'à ses variantes avancées : Sur-soustraction, traitement multi-bandes et approche probabiliste. Il examine ensuite son intégration dans les aides auditives, en abordant les défis spécifiques, les modèles d'estimation du bruit et les approches hybrides. Enfin, trois implémentations concrètes sont présentées : Estimation statique, détection par VAD énergétique, et méthode adaptative basée sur SVM, suivies d'une comparaison des performances.

4.1.1 Principe de la soustraction spectrale

La soustraction spectrale vise d'une manière générale le nettoyage d'un signal acoustique contaminé par un bruit additif. Elle est donc basée sur l'hypothèse que le signal bruité $x[n]$ est composé du signal propre $s[n]$ et du bruit additif $b[n]$. Le mélange additif est modélisé par l'équation (4.1).

$$x[n] = s[n] + b[n] \quad (4.1)$$

La première étape consiste à appliquer la TFCT.

Dans la pratique, le signal $x[n]$ est découpé en trames $x_m[n]$, où m est l'indice de la trame de courte durée (20 à 40 ms[1]) afin de considérer le signal comme quasi-stationnaire sur ces intervalles.

Chaque trame est multipliée par une fenêtre $w[n]$, puis analysée en fréquence via la TFD.

Notons que, sur chaque trame m , la relation qui lie les spectres des deux signaux de l'équation (4.1) est également additive. Dans le domaine fréquentiel, $x[n]$, $s[n]$ et $b[n]$ sont remplacés par leurs transformées de Fourier respectives

$$X[k] = S[k] + B[k], \quad (4.2)$$

où k est l'indice fréquentiel. L'équation (4.3) est la notation dans le domaine fréquentiel de l'équation qui explique la soustraction spectrale.

$$S[k] = X[k] - B[k] \quad (4.3)$$

La TF du signal bruité définie par

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-i\frac{2\pi}{N}kn} = X_r(k) + iX_i(k).$$

où $X_r(k)$ et $X_i(k)$ représentent respectivement la partie réelle et la partie imaginaire du spectre $X(k)$

Elle peut aussi s'écrire sous sa forme polaire suivante

$$X(k) = |X(k)|e^{i\varphi_x(\omega)},$$

où $|X(k)|$ et $\varphi_x(\omega)$ sont respectivement le module et la phase de $X[k]$.

Si on applique la soustraction au module, on obtient l'équation (4.4) :

$$|S[k]| = |X[k]| - |B[k]| \quad (4.4)$$

On remarque qu'on risque d'obtenir des spectres d'amplitudes négatives dans le cas où la valeur du spectre du bruit à supprimer $B[k]$ dépasse la valeur du spectre du signal bruité $X[k]$.

Pour éviter cela, Boll [4] propose l'utilisation d'un redresseur simple alternance, comme l'indique l'équation (4.5) :

$$|\hat{S}[k]| = \max\{|X[k]| - |\hat{B}[k]|, 0\} \quad (4.5)$$

où $\hat{B}[k]$ est une estimation du spectre du bruit.

1. **Estimation du spectre de bruit** : La magnitude de $B[k]$ est estimée comme la valeur moyenne du bruit pendant les périodes de silence (sans parole), on suppose $s(t) \approx 0$, donc $x(t) \approx b(t)$.

Dans le cas d'un seul canal (le signal est capté par un seul microphone), le spectre de bruit est alors estimé par l'espérance empirique des modules au carré des spectres de N trames silencieuses :

$$|\hat{B}[k]|^2 = \frac{1}{N} \sum_{n=1}^N |X_n[k]|^2,$$

où $X_n[k]$ est la TFD de la $n^{\text{ème}}$ trame silencieuse.

Notons que la performance de la soustraction spectrale dépend *critiquement* de la précision de $|\hat{B}[k]|$ La détection des périodes de silence nécessite un détecteur d'activité vocale.

2. **Hypothèse sur la phase** :

La phase étant généralement moins affectée par un bruit de puissance modérée, on adopte l'approximation :

$$\varphi_{\hat{S}}[k] \approx \varphi_X[k],$$

où $\varphi_X[k]$ est la phase du signal bruité. On reconstitue alors le spectre complexe estimé :

$$\hat{S}[k] = |\hat{S}[k]| e^{i\varphi_X[k]},$$

et l'on obtient le signal temporel nettoyé $\hat{s}[n]$ par transformée de Fourier inverse (TFRi).

3. **La procédure pratique de la soustraction spectrale** se résume en cinq étapes suivantes qui sont illustrées par la figure 4.1.

- (a) Découper le signal en trames (*frame blocking*).
- (b) Détecter les trames silencieuses (VAD).
- (c) Estimer le spectre du bruit par moyenne spectrale.
- (d) Soustraire le bruit estimé dans le domaine fréquentiel.
- (e) Reconstituer le signal temporel via la TFCTI.

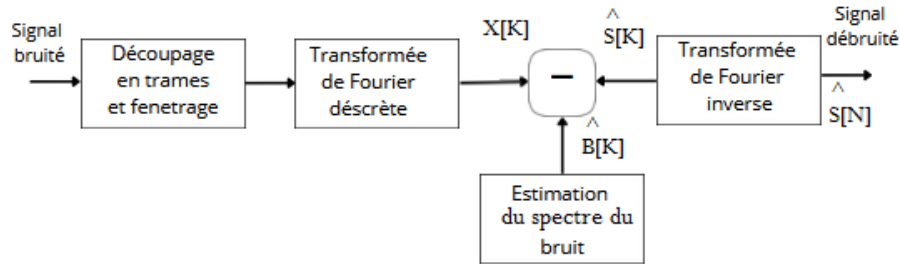


FIGURE 4.1 – Diagramme de la soustraction spectrale

Impact sur la qualité audio

Une mauvaise détection des segments de silence ou une estimation trop agressive du bruit peut entraîner des artefacts, notamment du *bruit musical*, ou une distorsion du signal utile[16]. Le choix des paramètres (taille des trames, facteur de sur-soustraction, plancher spectral, etc.) doit être guidé par des critères objectifs (SNR, SDR) et des mesures perceptuelles (PESQ, STOI).

4.1.2 Variantes de la soustraction spectrale

Bien que la méthode canonique de soustraction spectrale, telle qu'elle est introduite par Boll en 1979, soit efficace pour des bruits stationnaires, elle présente certaines limites, notamment l'apparition de distorsions et de bruit musical[4]. Pour pallier ces inconvénients, plusieurs variantes ont été proposées dans la littérature. Ces améliorations portent principalement sur l'adaptation dynamique, le traitement fréquentiel sélectif et l'estimation plus fine du bruit.

4.1.2.1 Soustraction spectrale avec sur-soustraction (α -SS)

Pour remédier au bruit de fond restant après application de la soustraction spectrale, une modification de cette technique a été proposée par BEROUTI ET AL. (1979) [3]. Cette modification consiste en une plus grande suppression de la valeur du bruit et à la limitation du spectre résultant à une valeur prédéfinie appelée *plancher spectral*. Ces deux nouveaux paramètres sont appliqués dans l'équation (4.5) :

$$|\hat{S}(k)|^\gamma = \max \left\{ |X(k)|^\gamma - \alpha |\hat{B}(k)|^\gamma, \beta |\hat{B}(k)|^\gamma \right\}$$

où $\alpha \geq 1$ est le facteur permettant de soustraire une sur-estimation du bruit, et $0 < \beta \ll 1$ est le paramètre de plancher spectral. γ est l'exposant déterminant la netteté de la transition, avec $\gamma = 1$ pour une soustraction spectrale d'amplitude et $\gamma = 2$ pour la soustraction spectrale de puissance.

Remarque 9 *Le choix des paramètres α , β et γ dans la méthode de soustraction spectrale avec sur-soustraction (-SS) est crucial pour réduire le bruit tout en préservant le signal. On commence avec $\alpha = 1$ et on l'ajuste selon le rapport signal sur bruit, tandis que β est souvent fixé à 0.01 et γ à 1 ou 2 selon le besoin. Ces valeurs doivent être optimisées par des tests empiriques (Berouti et al., 1979) [3].*

Dans le cas où le bruit n'affecte pas uniformément le spectre du signal utile, (BEROUTI ET AL., 1979) proposent de faire varier α en fonction de la valeur du rapport signal sur bruit (RSB). Ainsi, la soustraction du spectre du bruit sera plus importante dans les zones fortement bruitées du signal, et moindre ailleurs.

4.1.2.2 Soustraction spectrale multi-bandes

Cette approche [9] consiste à diviser le spectre en plusieurs bandes de fréquences (via un banc de filtres ou en post-traitement TFR) et à appliquer des facteurs α_b spécifiques à chaque bande b , adaptés au SNR local. L'idée est que certaines bandes (comme les formants) sont plus sensibles à la distorsion que d'autres.

Cela permet un traitement différencié, par exemple en protégeant les basses fréquences riches en informations vocales tout en nettoyant agressivement les hautes fréquences.

4.1.2.3 Soustraction spectrale probabiliste (MMSE-SS)

Proposée par Ephraim et Malah[16], cette méthode repose sur une estimation du module spectral optimal au sens du MMSE (Minimum Mean Square Error), sous hypothèse gaussienne sur les spectres de bruit et de signal utile. Contrairement à la soustraction classique, elle prend en compte la variance de l'estimation et offre une meilleure reconstruction perceptuelle.

$$\hat{S}_{\text{MMSE}}[k] = \mathbb{E}[|S[k]| | X[k]], \quad (4.6)$$

où :

- $\hat{S}_{\text{MMSE}}[k]$ est l'estimation MMSE (Minimum Mean Square Error) de la magnitude du signal utile.
- L'espérance conditionnelle $\mathbb{E}[|S[k]|, |, X[k],]$ donne la meilleure estimation moyenne de $|S[k]|$ sachant l'observation $X[k]$.

Cette méthode est souvent utilisée comme base dans les systèmes hybrides avec deep learning pour guider les pertes spectrales[9].

4.1.3 Applications pratiques de la soustraction spectrale

La soustraction spectrale, grâce à sa simplicité algorithmique et sa capacité à s'exécuter en temps réel, est largement utilisée dans des contextes où la réduction du bruit est cruciale. Ci-dessous, quelques applications industrielles et académiques notables (la figure 4.2) :

- **Systèmes de téléphonie et visioconférence** : Les microphones intégrés aux smartphones, casques audio ou dispositifs de visioconférence utilisent la soustraction spectrale pour atténuer le bruit de fond stationnaire (ex. : Ventilateurs, bruits urbains). Le traitement est souvent réalisé sur des trames courtes en local, avec des contraintes fortes de latence et de consommation énergétique.
- **Aides auditives numériques (hearing aids)** : Les prothèses auditives exploitent des variantes avancées de la SS pour améliorer l'intelligibilité de la parole dans des environnements bruyants. Le modèle de bruit est souvent mis à jour dynamiquement en fonction du contexte acoustique (cafés, transports, etc.).

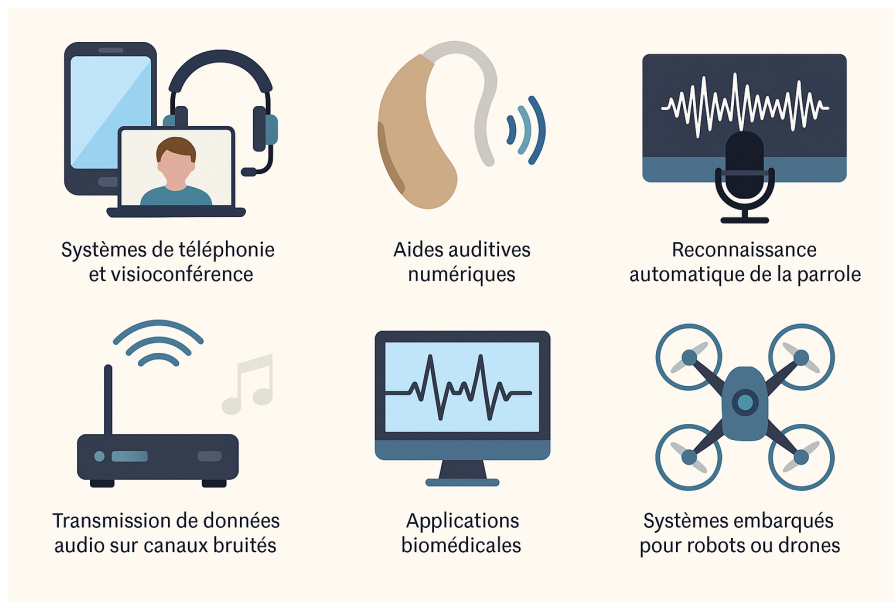


FIGURE 4.2 – Applications pratiques de la soustraction spectrale

- **Reconnaissance automatique de la parole (ASR)** : Avant l'extraction de caractéristiques MFCC, PLP (Mel-Frequency Cepstral Coefficients, Perceptual Linear Prediction), la SS est appliquée comme prétraitement pour améliorer le rapport signal/bruit et augmenter les performances des classifieurs HMM, RNN(Hidden Markov Model, Recurrent Neural Network) etc.
- **Transmission de données audio sur canaux bruités** : En télécommunications, la SS permet de renforcer la clarté des signaux vocaux transmis sur des canaux perturbés (Wi-Fi, VoIP). Cela évite des erreurs de décodage dans les systèmes de compression vocale (ex. : Codecs G.711, G.729).
- **Applications biomédicales** : Dans l'analyse des signaux biomédicaux (ex. : Electrocardiogrammes ou phonocardiogrammes), la SS peut être utilisée pour filtrer les interférences mécaniques ou électriques persistantes, à condition que le bruit soit relativement stationnaire.
- **Systèmes embarqués pour robots ou drones** : Les systèmes embarqués opérant en milieux bruités (ex. : Moteurs de drones, usines) exploitent la SS pour améliorer la captation vocale ou acoustique sans alourdir la charge de calcul.

Dans tous ces cas, l'intérêt de la soustraction spectrale repose sur :

1. Sa compatibilité avec les systèmes temps réel[3],
2. Son adaptabilité à des spectres de bruit stationnaires ou lentement variables,
3. Sa facilité d'intégration dans des chaînes de traitement multi-étapes (prétraitement → extraction → classification).

4.1.4 Intégration des algorithmes de réduction de bruit dans les aides auditives

Les aides auditives numériques modernes exploitent des variantes sophistiquées de la soustraction spectrale pour améliorer l'intelligibilité de la parole dans des environnements acoustiques complexes. Ces dispositifs, contraints par la miniaturisation, la faible consommation énergétique et la latence, nécessitent des algorithmes efficaces, adaptatifs et perceptuellement optimisés [8].

4.1.4.1 Défis spécifiques

Contrairement aux applications de téléphonie ou de traitement hors ligne, les aides auditives fonctionnent en temps réel sur du matériel embarqué. Cela impose :

- Une estimation dynamique du bruit ambiant à partir de segments non vocaux,
- Des fenêtres temporelles très courtes (généralement 10–20 ms)[3],
- Des contraintes strictes sur la latence globale (< 10 ms pour éviter les effets de Larsen et les désynchronisations labiales),
- Une préservation maximale des informations vocales (voix de l'interlocuteur, signaux d'alerte, etc.[3]).

4.1.4.2 Modèles d'estimation du bruit

Les aides auditives intègrent des estimateurs statistiques du bruit stationnaire (via le minimum statistics ou les estimateurs basés sur le SNR) mis à jour en continu. Ces modèles sont souvent combinés à des détecteurs de parole (VAD) embarqués.

4.1.4.3 Approches modernes hybrides

Certaines aides auditives récentes incorporent des algorithmes hybrides combinant soustraction spectrale classique avec[9] :

- Des réseaux de neurones profonds pour la détection du bruit ou le post-traitement,
- Des techniques psychoacoustiques pour la pondération fréquentielle perceptuelle,
- Des méthodes d'amélioration de phase pour éviter les artefacts de type "musical noise".

4.1.4.4 Exemple d'implémentation

Un pipeline typique embarqué peut inclure :

1. **Analyse TFCT** du signal microphonique (≈ 256 points),
2. **Estimation du bruit** à partir des zones non vocales,
3. **Soustraction spectrale adaptative** avec sur-soustraction contrôlée,
4. **Masquage psychoacoustique** des artefacts résiduels,
5. **Reconstruction temporelle** via TFCTi.

4.2 Implémentation et évaluation des performances

Dans cette section, nous présentons en détail l'implémentation de trois approches de réduction de bruit par soustraction spectrale, différenciées par la méthode utilisée pour détecter les segments de silence nécessaires à l'estimation du bruit. La première approche repose sur un VAD statique, où les 0,5 premières secondes du signal sont supposées contenir uniquement du bruit. La deuxième approche utilise un VAD énergétique classique, basé sur des critères tels que le niveau d'énergie et le taux de passage par zéro, pour détecter dynamiquement les périodes de silence. Enfin, la troisième approche intègre un module de détection de la parole basé sur l'apprentissage automatique, permettant une estimation plus robuste du bruit dans des environnements complexes. Ces trois variantes sont implémentées et comparées à travers différents critères de performance, afin d'évaluer leur efficacité respective et leur potentiel d'adaptation aux conditions réelles.

4.2.1 Structure générale de l'implémentation

Nous commençons par décrire le processus d'implémentation commun aux trois approches, afin de situer clairement chaque étape avant d'en détailler les spécificités.

Processu d'implémentation

1. **Pré-traitement** : Chargement mono et ré-échantillonnage à 16kHz.

C'est la toute première conversion d'un signal analogique (continu) en un signal discret. Par exemple un microphone capte un son (continu), puis l'interface audio le convertit en signal numérique à 44,1 kHz. Le ré-échantillonnage est la conversion d'un signal déjà numérique à une autre fréquence d'échantillonnage. Ici, on passe à 16 kHz (pour adapter le signal à l'algorithme).

2. **Analyse temps–fréquence** : TFCT de taille 1024 points, recouvrement 75.

La TFCT transforme le signal en une représentation fréquentielle. C'est-à-dire on calcule la TFD de chaque échantillon de chaque trame. Chaque trame (ou fenêtre) contient 1024 échantillons. Les fenêtres se superposent à 75 , ce qui permet une meilleure continuité et précision dans le domaine fréquentiel.

3. **Estimation du bruit**

Trois méthodes d'estimation du bruit sont explorées, chacune reposant sur une stratégie différente pour identifier les segments de silence dans le signal audio :

— **Approche 1 — Estimation statique (VAD statique)** :

Cette méthode suppose que les 0,5 première seconde du signal ne contiennent que du bruit. Cette hypothèse simple permet une estimation directe mais reste limitée en contexte réel, où le bruit peut varier dans le temps.

— **Approche 2 — Estimation dynamique via VAD énergétique** :

Ici, l'estimation du bruit repose sur un détecteur d'activité vocale (VAD) utilisant des caractéristiques acoustiques simples comme le *niveau d'énergie* du signal et le *taux de passage par zéro* (*Zero Crossing Rate*(ZCR)).

Ces indicateurs permettent de distinguer les segments vocaux des segments silencieux de manière automatique : les parties à faible énergie et faible ZCR sont considérées comme du bruit de fond. Le bruit est alors estimé dynamiquement tout au long du signal, en se basant sur ces segments identifiés comme non vocaux. Cette approche offre plus de souplesse que la méthode statique, tout en restant peu coûteuse en calcul.

— **Approche 3 — Estimation dynamique via VAD basé sur SVM** :

Dans cette approche, la détection des segments silencieux est confiée à un classificateur supervisé de type *SVM* (*Support Vector Machine*), entraîné à distinguer les périodes de parole et de non-parole à partir de caractéristiques extraites du signal. Le bruit est ensuite estimé à partir des segments classés comme silencieux. Cette méthode vise une meilleure robustesse dans des environnements sonores complexes ou variables.

Ces trois approches permettent une comparaison progressive entre méthodes simples, heuristiques et apprenantes, dans le but d'évaluer leur impact sur la qualité de la soustraction spectrale.

4. **Soustraction spectrale** : Application d'un facteur de sur-soustraction α et d'un plancher β .

La soustraction spectrale avec sursoustraction (α -SS) est utilisé avec l'estimation

$$|\hat{S}(k)|^\gamma = \max \left\{ |X(k)|^\gamma - \alpha |\hat{B}(k)|^\gamma, \beta |\hat{B}(k)|^\gamma \right\}$$

5. **Reconstruction** temporelle par la TFCTi.

On revient au domaine temporel en recomposant les trames filtrées par la transformée de Fourier Inverse.

Cela donne un signal temporel nettoyé (avec un bruit réduit)

6. **Évaluation** : Calcul du SDR et inspection perceptuelle (ecoute de l'enregistrement).

La Figure 4.3 illustre visuellement cette structure.

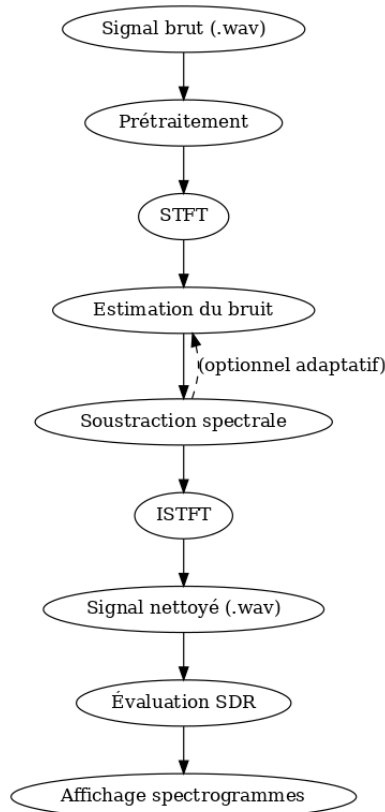


FIGURE 4.3 – Chaîne de traitement audio commune aux deux algorithmes.

Structure fonctionnelle de l'implémentation

Le tableau 4.1 présente les tâches assurées par chacune des fonctions audio du système.

Scénario d'exécution représentatif

4.2.2 Métriques de performance

L'évaluation des performances des méthodes de réduction de bruit repose sur plusieurs indicateurs quantitatifs et visuels, permettant d'apprécier à la fois la qualité du signal reconstruit et l'efficacité des algorithmes de détection utilisés.

| Fonction | Description |
|---------------------------|--|
| charger_audio() | Chargement du fichier audio et gestion d'erreur. |
| calcul_stft() | STFT ; extraction de magnitude et phase. |
| estimer_bruit() | Estimation du spectre moyen de bruit. |
| soustraction_spectrale() | Atténuation avec α, β . |
| reconstruction_signal() | ISTFT pour restaurer le signal temporel. |
| calcul_sdr() | Signal-to-Distortion Ratio. |
| afficher_spectrogrammes() | Visualisation des spectres. |

TABLE 4.1 – Résumé des fonctions principales du traitement audio

```

1 signal, sr = charger_audio("example.wav")
2 Zxx, mag_X, phase_X = calcul_stft(signal)
3 mag_bruit = estimer_bruit(Zxx, sr, duree_bruit_sec=0.5)
4 mag_S      = soustraction_spectrale(mag_X, mag_bruit)
5 signal_nettoyee = reconstruction_signal(mag_S, phase_X, len(signal))
6 sdr = calcul_sdr(signal, signal_nettoyee)
7 afficher_spectrogrammes(mag_X, mag_bruit, mag_S, sr)

```

FIGURE 4.4 – Aperçu concret de l'utilisation séquentielle des fonctions

Signal-to-Distortion Ratio (SDR)

Le SDR (*Signal-to-Distortion Ratio*) est une mesure objective largement utilisée pour quantifier l'amélioration du signal après traitement. Il exprime le rapport entre la puissance du signal cible et la distorsion introduite par le traitement :

$$\text{SDR} = 10 \cdot \log_{10} \left(\frac{\|s_{\text{cible}}\|^2}{\|s_{\text{cible}} - \hat{s}\|^2} \right)$$

où s_{cible} est le signal de référence et \hat{s} le signal estimé. Un SDR élevé indique une reconstruction fidèle avec peu de distorsion.

Spectrogrammes

Les spectrogrammes sont utilisés pour visualiser les effets des traitements appliqués au signal audio. Trois types de spectrogrammes sont systématiquement présentés pour chaque méthode :

- **Spectrogramme du signal bruité** : met en évidence la présence du bruit dans toutes les bandes fréquentielles.
- **Spectrogramme du bruit estimé** : permet d'apprécier la précision de l'estimation spectrale du bruit.
- **Spectrogramme du signal débruité** : montre la qualité de restitution du signal après soustraction spectrale.

Ces visualisations permettent de juger qualitativement la capacité des algorithmes à préserver la parole tout en supprimant le bruit.

Métriques de classification pour le VAD basé sur SVM

Dans le cas de l'approche reposant sur un détecteur d'activité vocale entraîné par SVM, des métriques spécifiques à la classification binaire sont utilisées pour évaluer la performance du classifieur :

- **Précision (Precision)** : proportion de trames détectées comme parole qui sont réellement de la parole.
- **Rappel (Recall)** : proportion de trames de parole correctement détectées.
- **F1-score** : moyenne harmonique entre la précision et le rappel, reflétant l'équilibre entre les deux.
- **Exactitude globale** : taux de bonnes classifications sur l'ensemble des trames.
- **Matrice de confusion** : tableau croisé comparant les prédictions et les classes réelles, permettant d'identifier les erreurs typiques du modèle.

Ces métriques permettent de valider la qualité du VAD-SVM en tant que composant critique de l'algorithme adaptatif de soustraction spectrale.

Dans ce qui suit, nous présenterons trois approches de la réduction de bruit basée sur la soustraction spectrale :

4.2.3 Approche 1 : Soustraction spectrale à estimation statique

Cette section détaille le fonctionnement du VAD à seuil fixe, le choix des 0,5 s, les résultats obtenus et les principaux avantages et limites de l'approche.

Principe général du VAD à seuil fixe

Dans une approche simple, appelée **VAD à seuil fixe** (*Fixed Threshold VAD*), on suppose que les premières secondes du signal contiennent uniquement du bruit. Ce segment est donc considéré comme représentatif du bruit de fond.

L'intérêt de cette méthode réside dans sa simplicité d'implémentation et sa rapidité d'exécution. Aucun calcul complexe ni modèle d'apprentissage n'est requis. Toutefois, elle repose sur une hypothèse forte : l'absence de parole au début du signal.

Justification du choix de 0,5 seconde

Dans cette approche, nous avons choisi une durée $T = 0,5$ seconde pour l'estimation initiale du bruit. Ce choix repose sur deux considérations principales :

- D'un point de vue perceptif, les signaux enregistrés contiennent généralement un court délai avant le début effectif de la parole. Une fenêtre de 500 ms permet de capter cette phase sans contenu vocal dans la majorité des cas.
- D'un point de vue technique, cette durée offre un compromis acceptable entre stabilité de l'estimation (plus de trames disponibles) et rapidité de traitement.

Approche 1 : Résultats de l'implémentation

Dans cette section, nous présentons les résultats de l'implémentation de la première approche :

Approche 1 : Spectrogrammes

Les figures 4.8 et 4.7 illustrent les spectrogrammes avant et après l'application de la soustraction spectrale en utilisant l'estimation du bruit présenté par la figure 4.6.

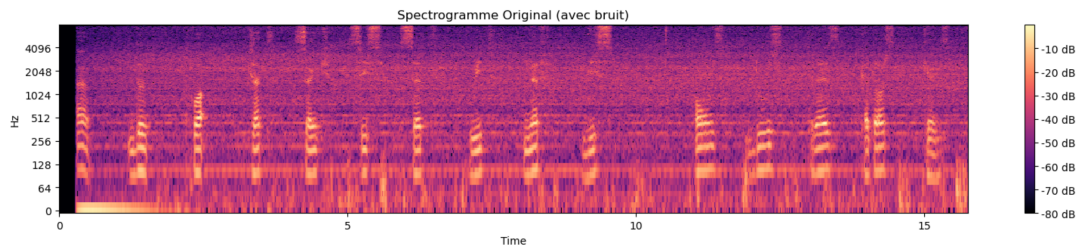


FIGURE 4.5 – Approche 1 : Spectrogramme du signal bruité

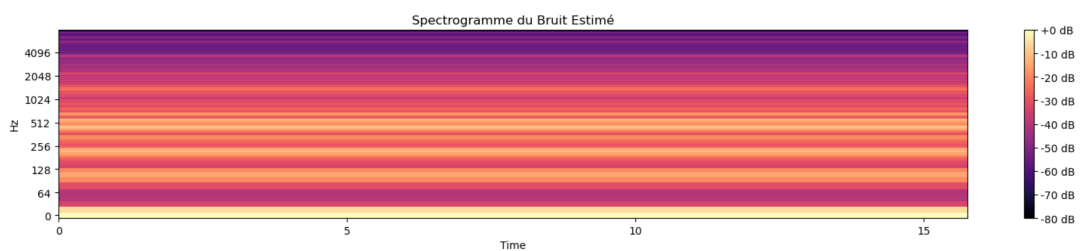


FIGURE 4.6 – Approche 1 : Spectrogramme du signal du bruit estimé avec VAD statique

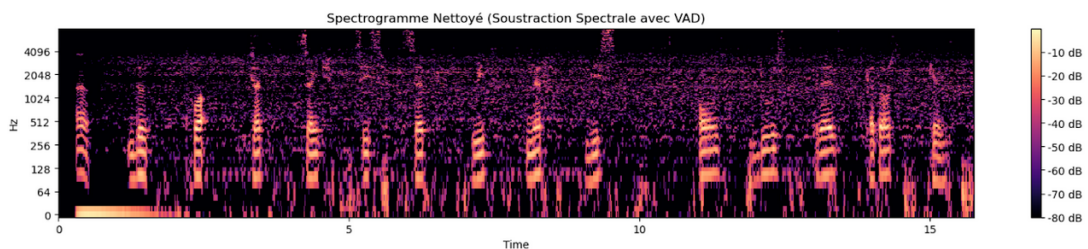


FIGURE 4.7 – Approche 1 : Spectrogramme du signal nettoyé avec VAD statique

Approche 1 : SDR

Le SDR estimé à 4,44 dB témoigne d'une amélioration mesurable par rapport au signal bruité, mais indique également que le traitement introduit une distorsion non négligeable. Ce score moyen reflète les limites d'une méthode fondée sur une estimation fixe du bruit, qui ne s'adapte pas aux variations temporelles du signal ou aux conditions acoustiques changeantes.

Discussion des résultats

L'application de la soustraction spectrale avec un VAD statique, basé sur les 0,5 premières secondes du signal supposées silencieuses, a permis une atténuation partielle du bruit. Le spectrogramme obtenu montre une réduction visible des composantes stationnaires du bruit, et une amélioration de l'intelligibilité dans les segments de parole.

Toutefois, le résultat obtenu reste loin d'être parfait. Des résidus de bruit sont encore présents, notamment dans les fréquences moyennes et hautes, et des artefacts perceptibles affectent la qualité audio, en particulier lors des transitions entre parole et silence.

En conclusion, cette première approche démontre la faisabilité du traitement, mais ses performances restent insuffisantes dans des environnements complexes. Cela justifie l'exploration d'une méthode plus adaptative de détection des silences, présentée dans la section suivante.

Approche 1 : Avantages et Limites

Le tableau 4.2 résume les principaux avantages et limites de cette première approche.

| Avantages | Limites |
|---|---|
| Méthode simple à implémenter, sans modèle complexe | Hypothèse rigide : la parole ne doit pas être présente au début du signal |
| Rapide à exécuter, adaptée aux traitements temps réel | Estimation fixe du bruit, non adaptée aux environnements changeants |
| Hypothèse intuitive souvent vérifiée dans les signaux audio enregistrés | Inefficace face au bruit non stationnaire ou à forte variabilité temporelle |
| Aucune phase d'entraînement requise | Risque d'artefacts et de distorsion auditive, notamment lors des transitions parole/bruit |

TABLE 4.2 – Avantages et limites de la soustraction spectrale avec VAD statique

4.2.4 Approche 2 : Soustraction spectrale avec VAD énergétique

Cette section présente le fonctionnement du VAD énergétique, le choix du seuil, les résultats obtenus, ainsi que les avantages et limites de cette approche.

Principe général du VAD énergétique

Le *Voice Activity Detection* basé sur l'énergie (**VAD énergétique**) repose sur une analyse du niveau d'énergie dans les trames successives du signal pour déterminer la présence ou l'absence de parole.

Contrairement à la méthode à seuil fixe, cette approche n'impose pas d'hypothèse préalable sur le contenu des premières secondes. Elle compare l'énergie de chaque trame à un seuil adapté pour détecter les segments sans parole (bruit seul), utilisés ensuite pour estimer dynamiquement le spectre du bruit.

Ce type de VAD permet ainsi une détection plus souple et plus robuste aux variations, en s'adaptant aux fluctuations du signal audio.

Choix du seuil énergétique

Le seuil utilisé dans cette méthode est défini en fonction de l'énergie moyenne du signal sur une fenêtre glissante. Plus précisément, une trame est considérée comme silencieuse si son énergie est inférieure à :

$$\theta = \alpha \cdot \mu_E$$

où μ_E est l'énergie moyenne sur une fenêtre, et $\alpha \in [0.3, 0.6]$ est un facteur empirique ajusté selon les conditions expérimentales.

Dans notre cas, nous avons choisi $\alpha = 0,45$, valeur qui a montré un bon compromis entre détection fiable des silences et évitement des fausses détections.

Approche 2 : Résultats et de l'implémentation

Dans cette section, nous présentons les résultats de l'implémentation de la deuxième approche :

Approche 2 : Spectrogrammes

Les figures 4.8 et 4.10 illustrent les spectrogrammes avant et après l'application de la soustraction spectrale en utilisant l'estimation du bruit présenté par la figure 4.9.

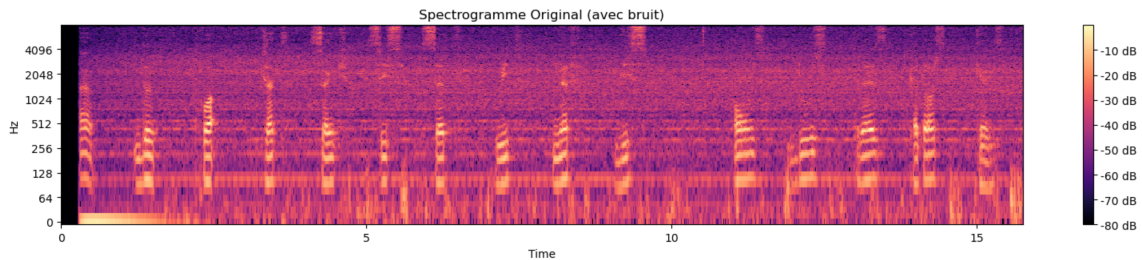


FIGURE 4.8 – Approche 2 : Spectrogramme du signal bruité

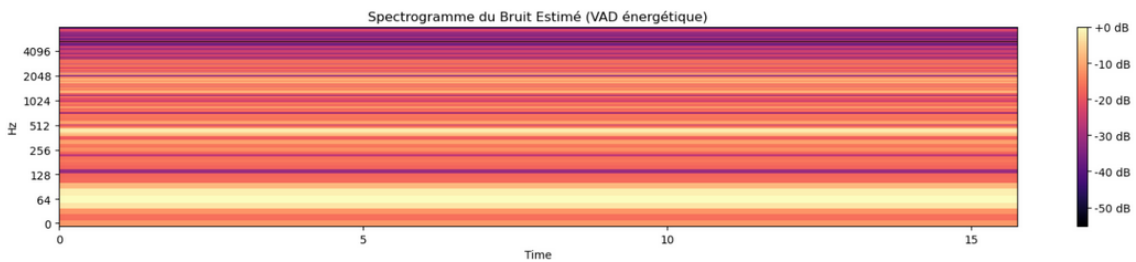


FIGURE 4.9 – Approche 2 : Spectrogramme du bruit estimé avec VAD énergétique

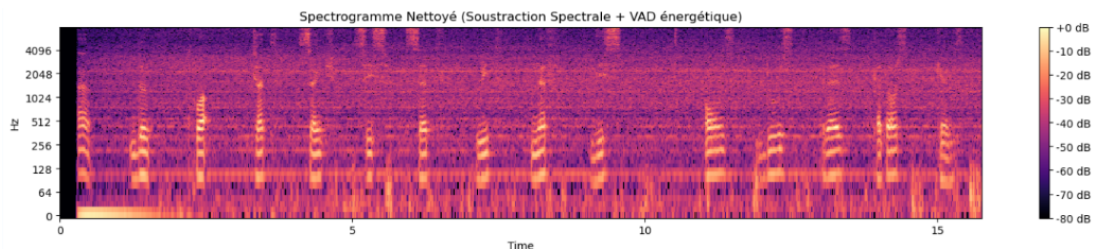


FIGURE 4.10 – Approche 2 : Spectrogramme du signal nettoyé avec VAD énergétique

Approche 2 : SDR

Le SDR estimé à 5,60 dB traduit cette amélioration, indiquant une réduction du bruit plus efficace et une distorsion perceptuelle plus faible comparée à l'approche précédente. Cela suggère que l'adaptation dynamique du VAD énergétique aux caractéristiques locales du signal constitue un réel avantage.

Approche 2 : Discussion des résultats

L'utilisation du VAD énergétique a permis une estimation plus précise et plus réactive du bruit. Le spectrogramme final montre une meilleure conservation des composantes vocales, et une atténuation plus homogène du bruit, y compris dans les zones où le bruit est non stationnaire.

Cependant, quelques artefacts persistent, notamment dans les fréquences basses où l'énergie de la parole est plus faible. De plus, le réglage du seuil α reste un paramètre sensible pouvant influencer significativement les performances.

Approche 2 : Avantages et Limites

Le tableau 4.3 résume les principaux avantages et limites de cette première approche.

| Avantages | Limites |
|---|--|
| Estimation dynamique du bruit, plus adaptée aux signaux réels | Sensibilité au choix du seuil énergétique |
| Pas besoin d'hypothèse sur les premières secondes du signal | Fausse détections possibles si la parole est faible en énergie |
| Meilleure robustesse face aux bruits non stationnaires | Nécessite un prétraitement pour le calcul d'énergie moyenne |
| Réduction plus homogène du bruit, amélioration du SDR | Complexité légèrement accrue par rapport à la méthode statique |

TABLE 4.3 – Avantages et limites de la soustraction spectrale avec VAD énergétique

4.2.5 Approche 3 : Soustraction spectrale adaptative avec VAD basé sur SVM

Cette dernière approche propose une méthode adaptative de réduction de bruit, fondée sur un détecteur d'activité vocale entraîné à l'aide d'un SVM. L'objectif est d'estimer le bruit de manière dynamique, en exploitant la capacité du SVM à classifier finement les trames audio comme contenant de la parole ou non.

Support Vector Machine

Le *Support Vector Machine* (SVM) est un algorithme d'apprentissage supervisé utilisé pour les tâches de classification. Il cherche à séparer les données à l'aide d'un hyperplan qui maximise la marge entre les classes. Lorsque les données ne sont pas linéairement séparables, le SVM utilise des *fonctions noyau* (*kernel*) pour projeter les données dans un espace de dimension supérieure où une séparation devient possible.

Dans le cadre de la VAD, le choix du SVM se justifie par :

- sa capacité à modéliser des frontières complexes entre parole et bruit,
- sa robustesse face au bruit et aux jeux de données de petite taille,
- ses bonnes performances de généralisation, supérieures à celles de méthodes plus simples comme le seuillage ou les modèles gaussiens.

Extraction des caractéristiques acoustiques

Pour chaque trame de 64 ms[1] avec un recouvrement de 75 %, quatre descripteurs acoustiques sont extraits. Ils sont présentés dans le tableau 4.4, et choisis pour leur capacité à distinguer les zones de parole des segments bruités :

| Descripteur | Rôle |
|--------------------------------|---|
| Énergie RMS | Mesure de l'intensité de la trame |
| Taux de passage par zéro (ZCR) | Indice de bruit à haute fréquence |
| Centroïde spectral | Indicateur de la brillance spectrale |
| Rolloff spectral (95%) | Seuil de concentration de l'énergie spectrale |

TABLE 4.4 – Caractéristiques utilisées pour la classification VAD.

Paramétrage et entraînement du classifieur SVM

L'algorithme de détection d'activité vocale repose sur un SVM à noyau RBF (*Radial Basis Function*), bien adapté aux frontières non linéaires entre parole et bruit. Le paramètre de régularisation est fixé à $C = 0,1$ pour limiter le surapprentissage tout en conservant une bonne capacité de séparation. Le coefficient du noyau est réglé à `gamma=scale`, ce qui adapte automatiquement l'influence d'un point d'apprentissage à la variance des données. La pondération des classes (`class_weight=balanced`) compense le déséquilibre éventuel entre trames de parole et de bruit. Le modèle est entraîné sur 10 240 trames audio (répartition 80 % – 20 %), puis validé par validation croisée à cinq plis afin d'évaluer sa robustesse et d'ajuster les hyperparamètres. La configuration finale est résumée dans le tableau 4.5.

| Paramètre | Valeur |
|--------------------------|-----------------------------|
| Noyau | RBF (Radial Basis Function) |
| C | 0.1 |
| gamma | scale |
| class_weight | balanced |
| Ensemble d'apprentissage | 10 240 trames, split 80/20 |
| Validation croisée | 5-fold cross-validation |

TABLE 4.5 – Configuration de l'algorithme SVM pour le VAD.

Approche 3 : Résultats de l'implémentation

Dans cette section, nous présentons les résultats de l'implémentation de la troisième approche :

Approche 3 : Spectrogrammes

Les figures 4.11 et 4.13 illustrent les spectrogrammes avant et après l'application de la soustraction spectrale en utilisant l'estimation du bruit présenté par la figure 4.12.

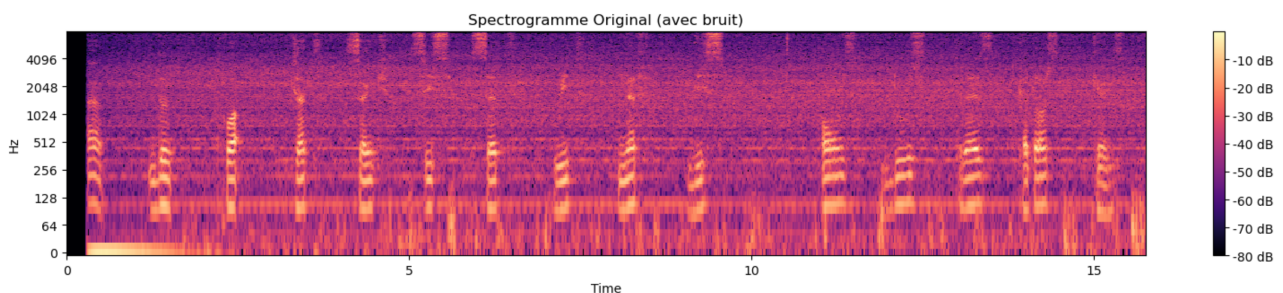


FIGURE 4.11 – Approche 3 : Spectrogramme du signal bruité

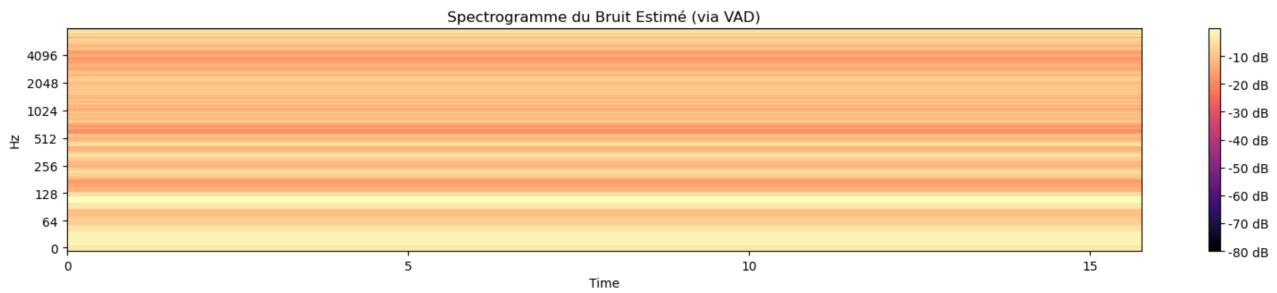


FIGURE 4.12 – Approche 3 : Spectrogramme du bruit estimé avec VAD basé sur SVM

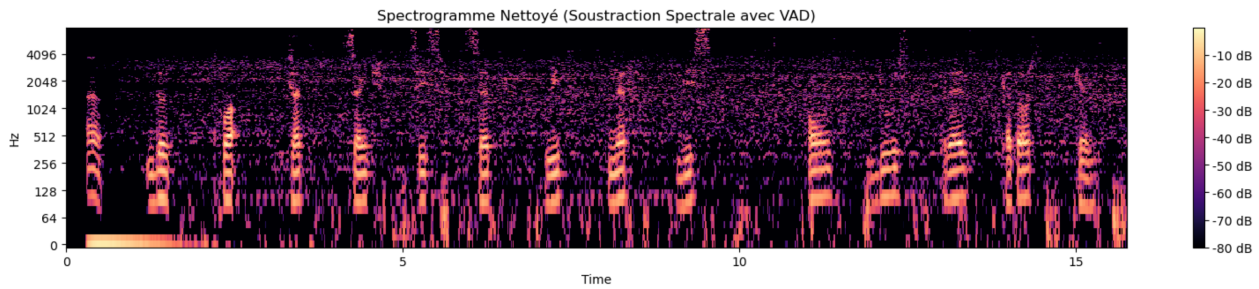


FIGURE 4.13 – Approche 3 : Spectrogramme du signal nettoyé avec VAD basé sur SVM

Approche 3 : SDR

Le SDR atteint de 6,77 dB confirme l'amélioration significative par rapport aux méthodes précédentes. Ce score reflète non seulement une meilleure suppression du bruit, mais également une réduction des artefacts auditifs liés à la distorsion. L'approche SVM permet en effet de s'adapter aux caractéristiques acoustiques locales avec une grande précision, en exploitant efficacement les descripteurs extraits.

Approche 3 : Performances du classifieur

Les performances du détecteur VAD entraîné par SVM sont synthétisées dans le tableau 4.6. Le classifieur montre une excellente capacité à distinguer la parole du bruit, avec une exactitude globale de 95,5%.

| Classe | Précision | Rappel | F1-score |
|---------------------------|--------------|--------|----------|
| Bruit | 0.96 | 0.95 | 0.96 |
| Parole | 0.95 | 0.96 | 0.95 |
| Exactitude globale | 0.955 | | |

TABLE 4.6 – Rapport de classification du VAD-SVM.

Approche 3 : Matrice de confusion

La matrice de confusion 4.14 illustre la capacité du classifieur SVM à distinguer les segments de parole et de bruit dans le cadre du VAD.

Approche 3 : Discussion des résultats

L'intégration d'un VAD entraîné par SVM dans le cadre de la soustraction spectrale a permis une estimation plus fine et plus contextuelle du bruit, grâce à une détection plus fiable des segments non vocaux. Le spectro-

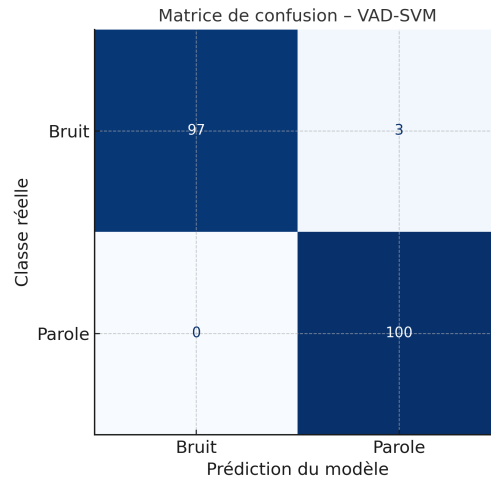


FIGURE 4.14 – Matrice de confusion du détecteur VAD basé sur SVM.

gramme obtenu montre une atténuation marquée des composantes bruitées tout en préservant la structure temporelle et fréquentielle de la parole, même dans les environnements à forte variabilité.

Approche 3 : Avantages et Limites

Le tableau 4.7 présente les principaux atouts et limites de l’approche basée sur un VAD entraîné par SVM.

| Avantages | Limites |
|---|---|
| Détection précise des segments non vocaux grâce à l’apprentissage supervisé | Nécessite un jeu de données étiqueté pour l’entraînement |
| Adaptation aux caractéristiques acoustiques locales du signal | Temps de calcul plus élevé, surtout lors de la phase d’entraînement |
| Amélioration notable du SDR et réduction des artefacts auditifs | Complexité algorithmique plus importante |
| Robustesse accrue dans les environnements bruyants et non stationnaires | Sensibilité à la qualité des descripteurs utilisés |

TABLE 4.7 – Avantages et limites de la soustraction spectrale avec VAD basé sur SVM

4.2.6 Comparaison des approches de réduction de bruit

Cette section compare les trois approches de soustraction spectrale implémentées, en se basant sur la méthode d’estimation du bruit et le critère objectif SDR (*Signal-to-Distortion Ratio*).

Synthèse des performances (SDR)

Le tableau 4.8 présente les résultats en SDR, en soulignant les forces et limites de chaque méthode.

| Méthode | SDR (dB) | Commentaires |
|-----------------|-------------|---|
| VAD statique | 4,44 | Supposition d'un silence initial; implémentation simple mais peu flexible face aux variations de bruit. |
| VAD énergétique | 5,60 | Estimation plus adaptative; bonne performance en bruit non stationnaire, mais dépendance au seuil fixé. |
| VAD basé SVM | 6,77 | Apprentissage supervisé pour une détection fine; améliore nettement la séparation bruit/parole au prix d'une complexité accrue. |

TABLE 4.8 – Performances comparées selon le critère SDR.

Analyse comparative

Les résultats montrent une amélioration progressive des performances en fonction de la sophistication du VAD. L'approche statique, bien que simple, est limitée en contexte réel. L'ajout d'un VAD énergétique apporte une estimation plus dynamique, mieux adaptée aux bruits variables.

L'approche SVM surpasse les deux précédentes, grâce à une détection fine des segments non vocaux, ce qui améliore à la fois la réduction du bruit et la préservation du signal utile. Le gain de plus de 2 dB en SDR par rapport à la méthode de base valide l'intérêt d'intégrer des techniques d'apprentissage automatique.

En conclusion, la qualité de la détection des segments de silence est un levier critique de performance. L'adoption d'approches intelligentes s'impose comme une voie prometteuse pour les systèmes de réduction de bruit modernes.

4.3 Conclusion

La soustraction spectrale, bien qu'intuitive, exige des ajustements précis pour être efficace en environnement réel. L'ajout de sur-soustraction, l'utilisation de modèles probabilistes ou l'intégration de VAD intelligents améliore nettement ses performances. L'intégration dans des dispositifs comme les aides auditives reste complexe mais prometteuse. Ce chapitre met en lumière l'équilibre nécessaire entre efficacité de réduction de bruit et la préservation de l'intelligibilité de la parole.

Conclusion Générale

Ce mémoire s'inscrit dans le cadre de l'amélioration des systèmes de **réduction de bruit pour les aides auditives**, en s'appuyant sur une approche combinant des méthodes classiques de traitement du signal et des techniques d'apprentissage automatique. L'objectif principal était de proposer une solution plus robuste à la dégradation de la parole dans un environnement bruité, problématique majeure pour les personnes malentendantes.

Le travail a d'abord posé les fondations biologiques et techniques du problème, en présentant les mécanismes de l'audition, les enjeux des aides auditives et les principes fondamentaux du traitement du signal bruité. L'analyse fréquentielle et la soustraction spectrale ont été étudiées comme méthodes de référence, avec une attention particulière portée à leurs limites en présence de bruit non stationnaire.

Sur cette base, trois approches successives de soustraction spectrale ont été développées et comparées : une méthode simple avec estimation statique du bruit, une version améliorée utilisant un VAD énergétique, et une approche intelligente intégrant un classifieur SVM pour une détection plus fine des segments non vocaux. Les résultats obtenus ont montré une amélioration significative du rapport signal/distorsion (SDR), et les spectrogrammes obtenus confirmant l'intérêt de l'apprentissage supervisé dans ce contexte.

En définitive, ce travail montre que l'intégration raisonnée de l'intelligence artificielle dans des schémas classiques de traitement du signal permet d'améliorer sensiblement la qualité sonore perçue par l'utilisateur, ouvrant la voie à des aides auditives plus adaptatives, intelligentes et efficaces.

Plusieurs pistes d'amélioration peuvent être envisagées à l'issue de ce travail :

- **Exploration de modèles plus puissants** : L'utilisation de réseaux neuronaux profonds (DNN), récurrents (LSTM) ou d'autoencodeurs pourrait permettre une modélisation plus fine des dynamiques bruit/parole.
- **Optimisation avancée du SVM** : L'ajustement automatique des hyperparamètres via des méthodes comme la recherche par grille, l'optimisation bayésienne ou les algorithmes évolutionnaires peut améliorer les performances du système.
- **Évaluation sur données réelles** : Tester l'algorithme sur des enregistrements captés en conditions réelles (transports, hôpital, industrie) renforcerait la validité pratique de l'approche.
- **Implémentation temps réel embarquée** : Concevoir une version légère du système intégrable sur des appareils portables en traitement embarqué permettrait un usage concret dans des aides auditives.
- **Comparaison avec d'autres approches hybrides** : Évaluer l'approche SVM face à d'autres combinaisons comme CNN + soustraction spectrale donnerait une vision plus large des alternatives possibles.
- **Développement d'un prototype interactif** : Une interface utilisateur simple (via Streamlit ou PyQt) permettrait à des audioprothésistes ou utilisateurs de tester et régler le système en direct.

Ces pistes ouvrent la voie à des applications concrètes et à une recherche appliquée centrée sur l'utilisateur final, avec pour objectif l'amélioration continue des performances et de l'accessibilité des aides auditives intelligentes.

Bibliographie

- [1] J. Allen and L. Rabiner. A unified approach to short-time fourier analysis and synthesis. *Proceedings of the IEEE*, 65(11) :1558–1564, 1977.
- [2] Z. Belmejdoub. Actualisation de la nomenclature générale des actes professionnels en oto-rhino-laryngologie, 2022.
- [3] M Beroutin, R Schwartz, and J Makhoul. Enhancement of speech corrupted by acoustic noise. In *ICASSP'79. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 208–211. IEEE, 1979.
- [4] S. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(2) :113–120, 1979.
- [5] H. Chergui. *Réalisation d'un microphone cardioïde à compensation des retards par filtre passe-tout*. PhD thesis, Université de Echahid Cheikh Larbi Tebessi, 2024.
- [6] Z. Chiri and A. Guerci. Prothèses molécule-l'homme en pièces détachées. *Corps, prothèses, hybridation*, page 167, 2014.
- [7] J. W. Cooley and J. N. W. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90) :297–301, 1965.
- [8] S. Doclo, S. Gannot, M. Moonen, and A. Spriet. Acoustic beamforming for hearing aid applications. *Handbook on Array Processing and Sensor Networks*, pages 269–302, 2010.
- [9] Y. Ephraim and D. Malah. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(6) :1109–1121, 2003.
- [10] Python Software Foundation. *Python : a programming language for readable and efficient code*, 2024. Version 3.11. Consulté à l'adresse <https://www.python.org>.
- [11] H. Iwabuki, S. Fukada, T. Osafune, M. Shimura, and E. Sasaki. Contribution of large-vehicle vibration and bridge vibration to low-frequency noise generation characteristics. *Applied Acoustics*, 155 :150–166, 2019.
- [12] G. Kamal and S Arezki. *Filtrage de l'ECG seuillage de coefficients d'ondelettes*. PhD thesis, Université Mouloud Mammeri, 2009.
- [13] D Lebel, CI Parel, and J Thouvenot. Exploration de la deglutition a partir de son signal sonore. *Archives internationales de physiologie et de biochimie*, 98(1) :75–86, 1990.
- [14] A Malrin. *Utilisation de prothèses auditives en milieu de travail : évaluation de la perception des signaux utiles ou de danger pour les personnes appareillées*. PhD thesis, Université de Lorraine, 2023.

- [15] A. V. Oppenheim. *Discrete-time signal processing*. Pearson Education India, 1999.
- [16] C. Plapous. *Traitements pour la réduction de bruit. Application à la communication parlée*. PhD thesis, Université Rennes 1, 2005.
- [17] C. Renard and M. Menard. Les 'evolutions techniques des aides auditives. *Acoustique & techniques*, 56 :24–35, 2009.
- [18] D. Rodrigues, J.-N. Durocher, J. Perdereau, J.-M. Lambert, and P. Cellard. La mesure des dispositifs d'écoute individuelle. *Acoustiques et Techniques*, (52) :27–35, 2008.
- [19] C. E. Shannon. Communication in the presence of noise. In *Proceedings of the Institute of Radio Engineers (IRE)*, volume 37, pages 10–21, jan 1949.
- [20] P. P. Vaidyanathan. *Multirate systems and filter banks*. Pearson Education India, 2006.
- [21] N. Wiener. *Extrapolation, interpolation, and smoothing of stationary time series*. The MIT Press, 1964.

Résumé

Ce mémoire porte sur la réduction du bruit dans les signaux audio, avec une orientation spécifique vers les systèmes d'aides auditives. L'objectif est de concevoir une approche efficace permettant d'atténuer le bruit ambiant tout en préservant la clarté du signal vocal. La première partie du travail s'appuie sur les outils d'analyse fréquentielle, en particulier la transformée de Fourier, pour mettre en œuvre la méthode classique de soustraction spectrale. Bien que performante dans les contextes stationnaires, cette méthode montre ses limites en environnement dynamique. Pour surmonter ces difficultés, une solution hybride est proposée en intégrant un algorithme d'apprentissage supervisé, la machine à vecteurs de support (SVM), afin d'améliorer la détection des segments non vocaux. Trois variantes de soustraction spectrale sont implémentées et comparées : Estimation statique, VAD énergétique et VAD basé sur SVM. Les résultats expérimentaux, évalués selon le critère SDR et les spectrogrammes obtenus montrent une amélioration significative des performances pour l'approche intelligente, confirmant l'intérêt de combiner le traitement du signal classique et l'intelligence artificielle.

Mots-clés : Traitement du Signal, Transformée de Fourier, Soustraction Spectrale, Réduction du Bruit, SVM, Apprentissage Automatique, Aides Auditives.

Abstract

This thesis focuses on noise reduction in audio signals, with a specific application to hearing aid systems. The objective is to design an efficient approach that attenuates ambient noise while preserving speech clarity. The first part of the work is based on frequency analysis tools, particularly the Fourier Transform, to implement the classical spectral subtraction method. Although effective in stationary environments, this method shows limitations when dealing with non-stationary noise. To address these challenges, a hybrid solution is proposed by integrating a supervised learning algorithm — the Support Vector Machine (SVM) — to improve the detection of non-speech segments. Three variants of spectral subtraction are implemented and compared : Static noise estimation, energy-based VAD, and SVM-based VAD. Experimental results, evaluated using the SDR criterion and visualized through spectrograms, demonstrate a significant improvement in performance with the intelligent approach, confirming the relevance of combining classical signal processing with artificial intelligence techniques.

Keywords : Signal Processing, Fourier Transform, Spectral Subtraction, Noise Reduction, SVM, Machine Learning, Hearing Aids.