

الجمهورية الجزائرية الديمقراطية الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE



UNIVERSITE MOHAMED SEDDIK BENYAHIA JIJEL

Faculté des sciences et de la technologie

Département d'Electronique

N° :...../2022

MEMOIRE DE FIN D'ETUDES

DOMAINE: Sciences et Technologies

FILIERE: Télécommunication

SPECIALITE: Systèmes des télécommunications

Thème

**Compression des signaux de parole par acquisition
comprimée pour les systèmes de télémédecine**

Présenté Par : Melle Bouchra MENIA

Encadré Par : Pr. Abdellah KACHA

Melle Loubna SAMER

Date de soutenance: 13/07/2022

Promotion : 2021 /2022

Remerciements

On remercie dieu le tout puissant de nous avoir donné la santé et la volonté d'entamer et de terminer ce mémoire.

Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de professeur, Mr Abdellah KACHA on le remercie pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur et sa disponibilité durant notre préparation de ce mémoire.

En second lieu, Mes sincères remerciements vont également aux membres du jury pour l'intérêt qu'ils portent à notre recherche en acceptant d'examiner notre travail et de l'enrichir de leurs propositions.

Enfin, nous tenons remerciements toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail.

Dédicaces

A nos chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien tout au long de nos études.

A nos chères sœurs, ..., pour leurs encouragements permanents, et leur soutien moral.

A nos chers frères, ..., pour leur appui et leur encouragement.

A nos amies, ..., pour leur soutien durant notre parcours universitaire.

Bouchra & Loubna

Table des matières

Remerciements.....	i
Dédicaces	ii
Table des matières.....	iii
Liste des figures	vi
Liste des tableaux	viii
Liste des abréviations.....	ix
Introduction générale.....	1

Chapitre I :

Notions de base sur la parole

I.1 Introduction.....	4
I.2 Production de la parole.....	4
I.2.1 L'appareilphonatoire.....	4
I.2.2 Le système articulatoire	6
I.3 Classification des sons selon la source.....	7
I.3.1 Les sons voisés.....	7
I.3.2 Les sons non voisés.....	8
I.4 Modèles source-filtre de production de la parole.....	8
I.5 Caractéristiques du signal de parole.....	10
I.5.1 Le phonème.....	10
I.5.2 La fréquence fondamentale.....	11
I.5.3 Les formants	11
I.5.4 Le spectre.....	11
I.5.5 L'énergie	12
I.6 Analyse de la parole pour l'évaluation des troubles de la voix.....	12
I.6.1 Les troubles de la parole et de la voix	12
I.6.2 Evaluation des troubles de la voix et de la parole.....	13
I.6.2.1 Evaluation subjective.....	13
I.6.2.2 Evaluation objective.....	14

I.7 Conclusion.....	16
---------------------	----

Chapitre II:

Systèmes de télémédecine

II.1 Introduction.....	18
II.2 Définition de la télémédecine.....	18
II.3 Les actes de la télémédecine	19
II.3.1 La téléexpertise.....	19
II.3.2 La téléconsultation.....	19
II.3.3 Téléassistance	19
II.3.4 La télésurveillance.....	20
II.3.5 Téléformation	20
II.4 La chaine de télémédecine	21
II.5 Avantages et inconvénients de la télémédecine.....	22
II.5.1 Avantages	22
II.5.2 Inconvénients.....	22
II.6 Le service de télémédecine pour les troubles de la parole et de la voix	22
II.6.1 Configuration du système.....	22
II.6.2 Codage et compression de la parole.....	24
II.6.3 Classification des algorithmes de compression	25
II.6.3.1 Compression sans perte.....	25
II.6.3.2 Compression avec perte.....	25
II.6.4 La technologie de compression MP3.....	25
II.6.4.1 Principe de fonctionnement.....	26
II.6.4.2 Algorithmes de codage.....	26
II.7 Conclusion	27

Chapitre III:

Acquisition comprimée

III.1 Introduction.....	28
III.2 Principe de l'acquisition comprimée	28
III.3 Parcimonie	30
III.4 Acquisition du signal	31
III.4.1 Incohérence	31
III.4.2 Mauvaise matrice de mesure	33
III.5 Algorithmes de récupération : théorie de l'optimisation.....	34
III.6 Exemple d'acquisition et de reconstruction.....	35
III.7 Conclusion	37

Chapitre IV:

Acquisition comprimée du signal de parole pour la télé-médecine

IV.1 Introduction	38
IV.2 Base des données	38
IV.3 Compression et reconstruction du signal de parole par acquisition comprimée	38
IV.3.1 Méthode	38
IV.3.2 Exemple d'illustration.....	40
IV.4 Résultats et discussion.....	42
IV.5 Conclusion.....	46
Conclusion générale	47
Références bibliographiques	48

Résumé

Liste des figures

Figure I.1: Anatomie de l'appareil phonatoire.....	5
Figure I.2: Le larynx.....	6
Figure I.3: Glotte ouverte et fermé.....	6
Figure I.4: Le processus oro-nasal.....	7
Figure I.5: Signal résiduel pour un son voisé [8].....	8
Figure I.6: Signal résiduel pour un son non voisé [8].....	8
Figure I.7: Modèle général du système de production de la parole.....	9
Figure II.1 : Exemple de systèmes multimédias en télécardiologie	19
Figure II.2 : Les types d'application de la télémédecine	21
Figure II.3 : Chaine de transmission de signal de parole pour les services de télémédecine.....	24
FigureIII.1:(a) Échantillonnage classique puis compression et (b) acquisition comprimée.....	28
Figure III.2 : Principe de l'acquisition comprimée.....	29
Figure III.3 : (a) Image mégapixel originale avec des valeurs de pixel dans la plage [0,255] et (b) ses coefficients de transformée en ondelettes (arrangés dans un ordre aléatoire pour une meilleure visibilité). (c) La reconstruction obtenue en mettant à zéro tous les coefficients de l'expansion en ondelettes sauf les 25 000 plus grands (les valeurs de pixel sont seuillées dans la plage [0, 255])......	31
Figure III.4 : Exemple d'une mauvaise matrice de mesure.....	34
Figure III.5 : En haut : Échantillons prélevés aléatoirement du signal original. En bas : la transformée en cosinus discrète inverse du signal.....	36
Figure III.6: Solution de l'équation $Ax = b$ obtenue par minimisation l1.....	36
Figure III.7 : Solution de l'équation $Ay = b$ obtenue par minimisation l2.....	37
Figure IV.1: Schéma bloc de la méthode de compression/reconstruction du signal de parole par acquisition comprimée.....	40
Figure IV.2: Signal reconstruit par acquisition comprimée (en rouge) superposé au signal original (en bleu). (a) score 0. (b) score 2.....	41
Figure IV.3: Corrélation entre les scores de l'évaluation auditive et les valeurs du jitter calculé à partir du signal reconstruit par acquisition comprimée et à partir du signal original et du signal MP3.....	43

Figure IV.4: Corrélation entre les scores de l'évaluation auditive et les valeurs du shimmer calculé à partir du signal reconstruit par acquisition comprimée et à partir du signal original et du signal MP3.....	44
Figure IV.5: Corrélation entre les scores de l'évaluation auditive et les valeurs du HNR calculé à partir du signal reconstruit par acquisition comprimée et à partir du signal original et du signal MP3.....	45
Figure IV.6: Corrélation entre les scores de l'évaluation auditive et les valeurs du SDR calculé à partir du signal reconstruit par acquisition comprimée et à partir du signal original et du signal MP3.....	46

Liste des tableaux

Tableau VI.1 : Valeurs des indices acoustiques jitter, shimmer et HNR calculés à partir du signal reconstruit et du signal original pour un locuteur normophonique (score 0) et un locuteur dysphonique (score 2).....42

Liste des abréviations

AC	Acquisition Comprimée
ADSL	Asymmetric Digital Subscriber Line
APQ3	Three-point Amplitude Perturbation Quotient
APQ5	Five-point Amplitude Perturbation Quotient
MP3	MPEG audio layer 3
CS	Compressed Sensing
dB	Décibel
DCT	Discrete Cosine Transform
D.C.E	Data Communication Equipment
D.T.E	Data Terminal Equipment
GRBASI	Grade, Roughness, Breathiness, Asthenia, Strain, Instability
HNR	Harmonic to Noise Ratio
Jitta	Jitter Absolut
Jitt	Jitter Locale ou Relative
JPEG	Joint Photographic Experts Group
MPEG	Motion Picture Experts Group
PPQ5	five-point Period Perturbation Quotient
RAP	Relative Average Perturbation
RIP	Restricted Isometry Property
RTC	Réseau Téléphonique Commuté
SDR	Signal-to-Dysperiodicity Ratio
SNR	Signal to Noise Ratio
ShdB	Shimmer Absolu ou Local
Shim	Shimmer Local
TAP	Traitement Automatique de la Parole

Introduction générale

Introduction générale

La télémédecine est définie comme une forme de pratique médicale à distance utilisant les technologies de l'information et de la communication [1]. La demande croissante des services de soins de santé a incité les chercheurs à rechercher de nouvelles façons d'organiser et de fournir les services soins de santé. La télémédecine, en tant que nouveau domaine de recherche, exploite les systèmes multimédias comme moyen d'augmenter la disponibilité des soins pour les patients dans le but d'économiser de temps et l'argent. La télémédecine offre un moyen pour les patients d'être examinés et traités, tandis que le fournisseur de soins de santé et le patient se trouve à différents endroits physiques. De même, la télémédecine a été proposée comme solution à la problématique des « déserts médicaux » afin d'améliorer l'accès aux soins à la population des zones rurales [1].

La télémédecine offre plusieurs services tels que la télécardiologie, la téléradiologie, la téléchirurgie, etc. Dans le contexte des applications cliniques de la parole, la télémédecine peut être utilisée pour la détection des pathologies laryngées et le suivi de l'évolution de la voix du patient durant les périodes de rééducation et de thérapie.

Grâce aux développements technologiques importantes, les futurs hôpitaux fourniront des services de soins de santé aux patients du monde entier en utilisant des systèmes multimédias et des signaux pouvant être acquis à distance. La transmission du signal et d'image, leur stockage et traitement sont les principaux composants d'un système de télémédecine.

Le stockage de signaux et d'images de haute qualité nécessite un espace mémoire important. Par conséquent, la transmission de tels signaux nécessite un réseau avec une large bande passante. La réduction de la bande passante et de l'espace mémoire requis, tout en gardant une qualité audio est réalisée par des algorithmes de compression. Les algorithmes de compression ont connu un progrès significatif. Cependant, il existe un besoin croissant de transférer une grande quantité de données sur le réseau de communication. Par conséquent, les algorithmes de compression ont un impact significatif sur les performances des systèmes de télémédecine.

Les applications de télémédecine à l'analyse et l'évaluation des troubles de la voix sont basées sur les technologies existantes pour la compression et la transmission du signal de parole signal développé pour les communications cellulaires et Internet. Il est important donc de connaître l'effet des algorithmes de compression sur la parole pathologique et la

Introduction générale

façon dont ils affectent les indices acoustiques utilisées comme descripteurs des troubles de la voix.

Les effets du taux d'échantillonnage et de compression sur l'intelligibilité de la parole et l'identification de voyelle produite par les personnes dysarthriques a été considéré dans [2]. Les résultats de l'étude réalisée sur deux locuteurs dysarthriques suggèrent que les changements dans l'intelligibilité peuvent être un artefact des algorithmes de compression utilisés pour transmettre la parole du client au praticien.

Dans ce mémoire, on se propose d'utiliser l'acquisition comprimée comme alternative aux algorithmes conventionnelles pour la compression des signaux de parole dans les applications de télémédecine pour le suivi à distance des personnes atteintes de troubles de la voix. L'acquisition comprimée est une méthode d'acquisition et de compression simultanées des signaux. Elle permet la reconstruction des signaux ayant une représentation parcimonieuse dans certaines bases en utilisant un nombre d'échantillons beaucoup plus faible que celui déterminé par le théorème de Shannon-Nyquist.

L'acquisition comprimée est utilisée comme un algorithme de compression du signal de parole produite par des locuteurs pathologiques. Le signal compressé reçu par le praticien est reconstruit pour calculer des indices acoustiques servant comme descripteurs de la qualité de la voix. L'effet de compression est investigué en comparant les indices acoustiques obtenus à partir du signal reconstruit à ceux calculés en utilisant le signal original et le signal compressé MP3.

La mémoire est organisée comme suit :

Le premier chapitre présente des notions de base sur la production de la parole. On rappelle brièvement le mécanisme de production de la parole puis on présente des caractéristiques de signal de parole ainsi que quelques indices acoustiques.

Le deuxième chapitre présente le système de télémédecine et l'analyse du signal de parole produite par des locuteurs pathologiques.

Le troisième chapitre présente la théorie de l'acquisition comprimée et ses différentes applications.

Le quatrième chapitre présente les résultats expérimentaux obtenus en utilisant une base de données incluant des signaux de parole produite par des locuteurs sains et

Introduction générale

pathologiques. Les performances de la méthode proposée seront comparées à celles de compression MP3 et le signal original.

Enfin, on termine par une conclusion qui résume les résultats obtenus.

chapitre I :

Notions de base sur la parole

I.1 Introduction

Le processus de production de la parole est un mécanisme très complexe, qui repose sur l'interaction entre le système nerveux et le système physiologique. La parole commence par l'activité neuronale. Le système nerveux joue un rôle important, il est donc responsable de la génération et du décryptage des signaux de parole. Après les pensées et la volonté de parler, le cerveau guide les opérations liées à l'activation des organes vocaux. La fonction de ces organes est de nature physiologique [3,4].

L'analyse acoustique pour l'évaluation des troubles de la voix est basée sur l'extraction d'indices acoustiques à partir du signal de parole en utilisant les techniques de traitement du signal. Dans un système de télémédecine, le signal est transmis après compression. Par conséquent, la compression du signal est un facteur important dans la détermination de la performance du système.

Ce chapitre est consacré aux notions de base nécessaire à la compréhension du mécanisme de production de la parole. On commence par une présentation du modèle de l'appareil phonatoire, ainsi que son modèle source-filtre. Ensuite, on présente les caractéristiques de la parole et les méthodes d'analyse pour l'évaluation des troubles de la voix.

I.2 Production de la parole

Les sons de la parole se produisent via le mouvement de nombreux muscles et organes de phonation. La coordination de ces organes est contrôlée par le système nerveux central [5].

I.2.1 L'appareil phonatoire

L'appareil phonatoire Figure I.1 se compose d'un ensemble d'organes qui permettent de produire la parole.

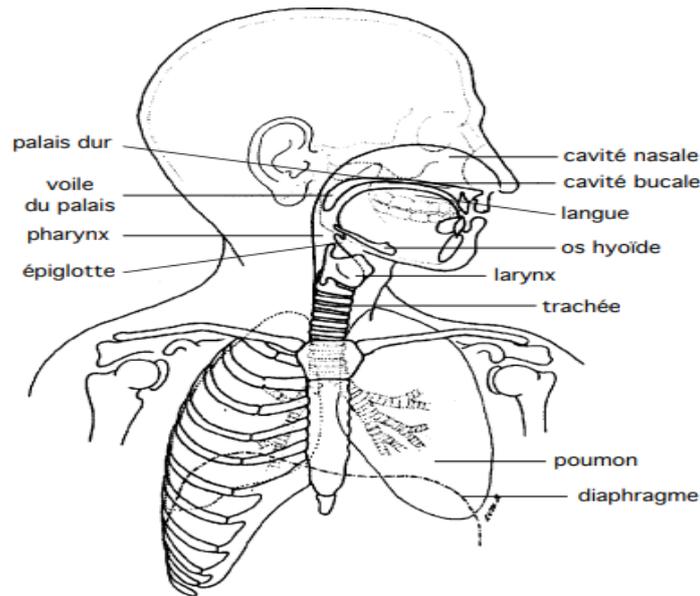


Figure I.1 : Anatomie de l'appareil phonatoire.

La parole est produite par un flux d'air provenant des poumons, qui traverse la trachée et les cavités buccale et nasale. Elle implique trois processus : l'initiation, la phonation, le processus oro-nasal et l'articulation.

Le processus d'initiation se produit lorsque l'air est expulsé des poumons. En anglais, les sons de la parole sont le résultat d'un « flux d'air pulmonaire » cependant, ceci n'est pas le cas dans toutes les langues.

Le processus de phonation se produit au niveau du larynx. Le larynx a deux plis horizontaux formés de tissus dans le passage de l'air; ce sont les cordes vocales comme montre la Figure I.2, et le trou entre ces plis s'appelle la glotte.

Lorsque la glotte est fermée, aucun air ne peut passer mais elle peut avoir une ouverture étroite qui peut faire vibrer les cordes vocales produisant le "son vocal". Enfin, elle peut être grand ouverte, comme dans la respiration normale ce qui réduit la vibration des cordes vocales et permet la production des sons non voisés comme montré dans la Figure I.3.

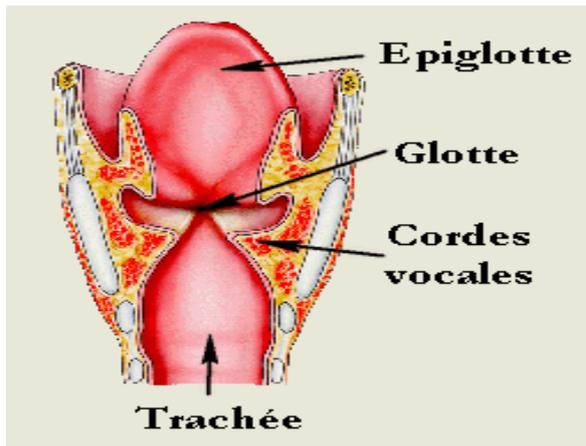


Figure I.2 : Le larynx.

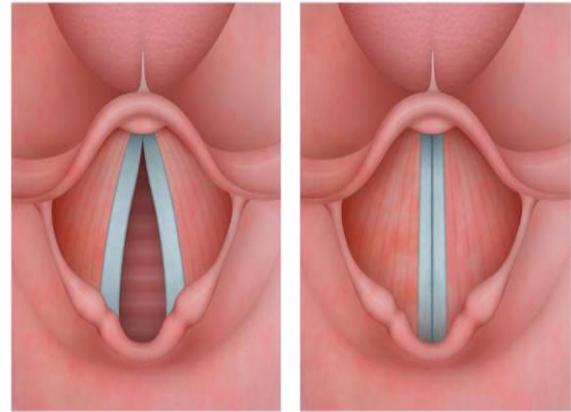


Figure I.3 : Glotte ouverte et fermé.

Après avoir traversé le larynx et le pharynx, l'air peut pénétrer dans la cavité nasale ou buccale. Le voile du palais est la partie responsable de cette sélection, comme on peut le voir sur Figure I.4 Grâce au processus oro-nasal, nous pouvons différencier entre les consonnes nasales (/m/, /n/, /N/) et les autres sons.

I.2.2 Le système articulatoire

Le système d'articulation se situe dans la bouche entre les cordes vocales et les lèvres, c'est le système par lequel nous pouvons différencier la plupart des sons. Il comprend un ensemble d'organes mobiles, le voile du palais, la mâchoire inférieure (ou mandibule), la langue et les lèvres. Les mouvements de la mâchoire inférieure contribuent largement aux variations de volume de la bouche. La langue, reliée par sa base à l'os hyoïde, est extrêmement mobile car commandée par dix-sept muscles (huit pairs et un impair) ; elle prend appui sur différents points du conduit pharyngo-buccal pour articuler les phonèmes. Ces points d'articulation sont :

- Les lèvres (articulations labiales ou bilabiales).
- Les dents (articulations dentales).
- Les alvéoles (articulations alvéolaires).
- Le palais dur, ou partie osseuse de la voûte (articulations palatales).
- Le voile du palais ou « palais mou » (articulations vélaires).
- la luvette (articulations uvulaires).
- le pharynx (articulations pharyngales).
- la glotte (articulations glottales).

Les lèvres sont également très mouvantes grâce à leurs treize paires de muscles. Le voile du palais, prolongement membraneux du palais dur ou osseux, s'abaisse sous l'action des pharyngo staphylins pour ouvrir la communication entre les voies respiratoires et les fosses nasales : les phonèmes émis sont alors « nasalisés ». (par exemple o, a, i, deviennent on, an, in) ; quand le voile se relève, sous l'action des muscles péri-staphylins internes, le conduit pharyngo-nasal est fermé et l'air expiré passe seulement par la bouche comme montré par la Figure I.4 [6].

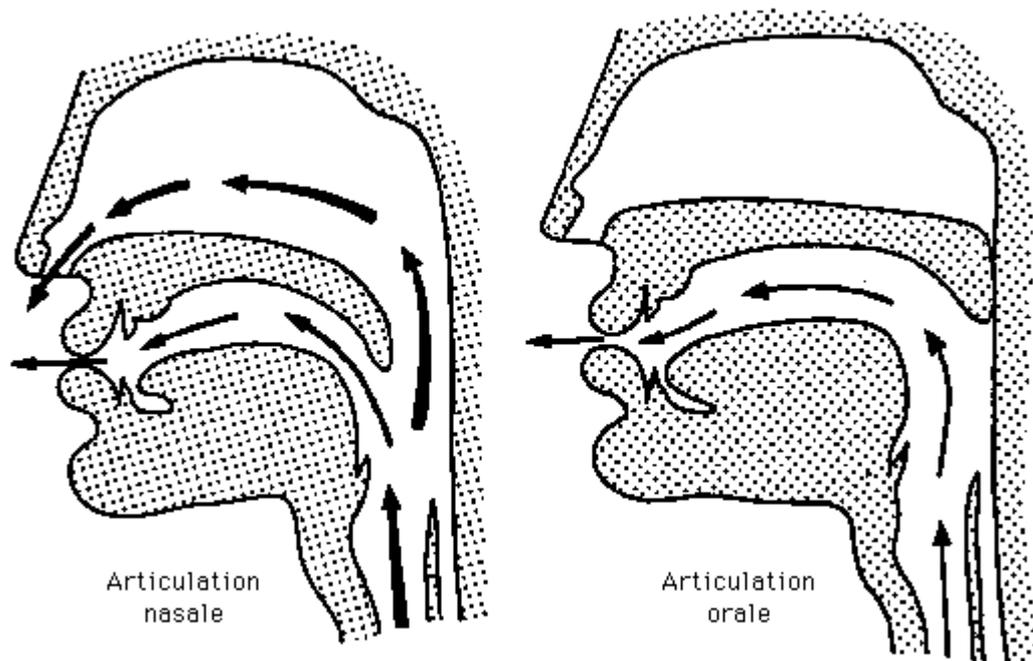


Figure I.4 : Le processus oro-nasal.

I.3 Classification des sons selon la source

Les sons peuvent être classés selon l'état des cordes vocales en sons voisés et sons non voisés.

I.3.1 Les sons voisés

Les sons voisés tels que les voyelles par exemple, sont produits par le passage de l'air provenant des poumons à travers la trachée, qui met en vibration les cordes vocales [7]. Pour la parole voisée, l'excitation possède un caractère périodique et des propriétés particulières dues à la forme de l'onde de débit glottique. Les sons voisés sont généralement quasi-périodiques. Ce type de sons représente la majorité du temps de phonation, et il est caractérisé en général par une énergie élevée en basse fréquence avec

environ un formant par kHz de bande passante, et dont seuls les trois ou quatre premiers contribuent de façon importante à l'information linguistique [7].

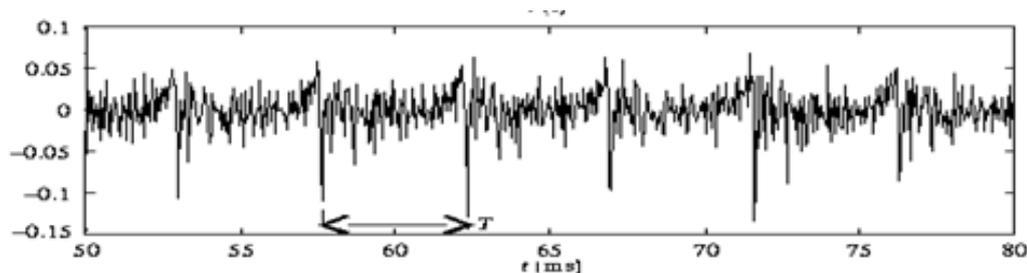


Figure I.5 : Signal résiduel pour un son voisé [8].

I.3.2 Les sons non voisés

Les sons non voisés présentent une structure apériodique, les cordes vocales sont écartées et ne vibrent pas. L'énergie de ce type de sons est concentrée dans les hautes fréquences et correspond à du bruit [7].

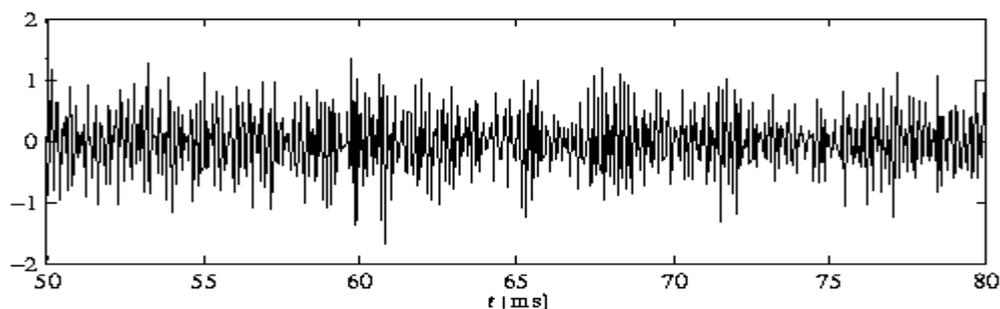


Figure I.6 : Signal résiduel pour un son non voisé [8].

I.4 Modèles source-filtre de production de la parole

La Figure I.7 représente le modèle discret source-filtre utilisé pour la modélisation du système de production de la parole [9]. Le processus de production de la parole peut être modélisé comme une opération de filtrage linéaire. Le filtre de fonction transfert $H(z)$ modélise le conduit vocal et les articulateurs, le filtre de fonction de transfert $R(z)$ simule le rayonnement aux lèvres. Pendant les segments de parole voisée, une estimation de la période fondamentale T_0 est utilisée par l'excitation pour la génération d'une suite d'impulsions qui excitent le filtre d'impulsion glottique $G(z)$. Pendant les segments de parole non voisée, un bruit blanc caractérisé par un spectre plat modélise l'excitation.

La fonction de transfert du conduit vocal peut être exprimée par [9] :

$$H(z) = \frac{G_0}{1 + \sum_{k=1}^N b_k z^{-1}} = \frac{G_0}{\prod_{k=1}^p (1 - p_k z^{-1})} \quad (I.1)$$

où G_0 représente un gain et $p_k, k = 1, \dots$, sont les pôles de la fonction transfert.

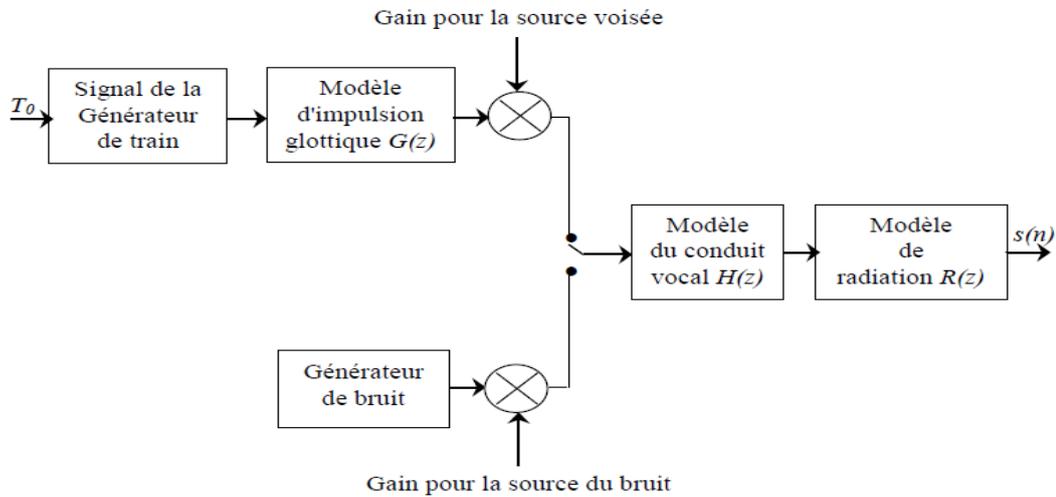


Figure I.7 : Modèle général du système de production de la parole.

Bloc 1 : Générateur de train d’impulsions

Dans le cas de la parole voisée, la source glottique est représentée par un modèle qui dépend de l’application, à condition que ce modèle soit tout pôle. Un modèle de la source glottique à deux pôles est donné par [9] :

$$G(n) = (\alpha^n - \beta^n)u(n), \quad \beta < \alpha < 1, \quad \alpha \approx 1 \quad (I.2)$$

$u(n)$ représente la séquence échelon unité.

La fonction de transfert modélisant le filtre de rayonnement aux lèvres est donnée par :

$$R(z) = 1 - z_0 z^{-1} \quad z_0 \approx 1, z_0 < 1 \quad (I.3)$$

Dans le cas de la parole non voisée, la sortie $S(z)$ du modèle est définie par:

$$S(z) = E(z)H(z)R(z) \quad (I.4)$$

où $E(z)$ représente la transformée en z d’une réalisation du bruit blanc.

Dans le cas de la parole voisée, la sortie $S(z)$ est donnée par :

$$S(z) = E(z)G(z)H(z)R(z) \quad (\text{I.5})$$

où $E(z)$ représente la transformée en z du train d'impulsions caractérisé par une période égale à la période fondamentale T_0 du signal de parole.

La fonction de transfert du système globale $\theta(z)$ du modèle peut être exprimée par :

$$\theta(z) = \begin{cases} H(z)R(z), & \text{pour les sons non voisés} \\ G(z)H(z)R(z), & \text{pour les sons voisés} \end{cases} \quad (\text{I.6})$$

I.5 Caractéristiques du signal de parole

La parole est un signal réel, continu, d'énergie finie et non stationnaire à long terme. Sa structure est complexe et variable avec le temps [10]. Sa composition est :

- Pseudopériodique pour les sons voisés.
- Aléatoire pour les sons fricatifs.
- Impulsionnelle avec la phase explosive pour les sons occlusifs.

Il existe différentes méthodes de représentation du signal. Certaines ont été spécifiquement développées pour l'analyse ou le codage [10].

I.5.1 Le phonème

Le phonème [11] est la plus petite unité présente dans la parole et susceptible par sa présence de changer la signification d'un mot. Le nombre de phonèmes est toujours très limité, en générale il est inférieur à 50.

Les phonèmes peuvent être rangés en catégories [12] selon des traits distinctifs qui indiquent une similitude au niveau articulaire, acoustique ou perceptif. On retrouve les voyelles et les consonnes. Les voyelles sont caractérisées par la vibration des cordes vocales, lieu de l'articulation et la stabilité des articulations produisant des sons tenus pendant un certain laps de temps, les voyelles peuvent être rangées selon :

- La nasalité.
- L'ouverture du conduit vocal.
- La position de la constriction du conduit vocal.
- L'arrondissement des lèvres.

Les consonnes se prononcent avec un rétrécissement du passage de l'air et sont classées selon:

- Le voisement (selon que les cordes vocales vibrent ou non à leur passage).
- Le mode d'articulation (occlusif, nasal, fricatif).
- Le lieu d'articulation (labiale, dentale, palatale).

I.5.2 La fréquence fondamentale

Elle est appelée aussi pitch et elle représente la fréquence de vibration des cordes vocales. Elle varie :

- De 150 Hz à 450 Hz pour une voix féminine.
- De 80 Hz à 200 Hz pour une voix masculine.
- De 200 Hz à 600 Hz pour une voix d'enfant.

I.5.3 Les formants

Le spectre du signal vocal résultant de l'action de la source sur le conduit vocal présente des maximums et des minimums qui correspondent aux résonances et aux antirésonances du conduit vocal, appelés formants et anti-formants. Du point de vue perceptif, seul les trois premiers formants jouent un rôle essentiel pour caractériser le spectre vocal [7]. On peut caractériser toute voyelle en n'utilisant que ses trois premiers formants. En général la fréquence du premier formant varie de 200 à 900 Hz, celle du second de 500 à 2500 Hz et le troisième se situe entre 1500 et 3500 Hz. Des formants d'ordre supérieur existent même si leur rôle sur le plan perceptif est limité, ils contribuent à caractériser la voix.

I.5.4 Le spectre

L'enveloppe spectrale ou spectre représente l'intensité du signal acoustique en fonction de la fréquence, elle est généralement obtenue par une analyse de Fourier à court terme. La quasi stationnarité du signal de parole permet de mettre en œuvre des méthodes efficaces d'analyse et de modélisation utilisées pour le traitement à court terme du signal acoustique sur des fenêtres de durée généralement comprise entre 20 ms et 30 ms appelées trames, avec un recouvrement entre ces fenêtres qui assure la continuité temporelle des caractéristiques du signal lors de son analyse.

I.5.5 L'énergie

L'énergie est un paramètre essentiel due à la force du flux pulmonaire mais elle peut être modifiée par l'ouverture des lèvres créant la résonance. Elle modélise l'intensité avec laquelle le locuteur produit le signal acoustique. L'énergie d'un signal $s(n)$ sur une fenêtre de longueur N est définie par [13] :

$$E = \sum_{n=0}^{N-1} |s^2(n)| \quad (\text{I.7})$$

I.6 Analyse de la parole pour l'évaluation des troubles de la voix

L'analyse du signal vocal fournit une méthode préférée pour l'évaluation clinique de la qualité de la voix. Le but de l'analyse du signal acoustique est d'extraire des indices pertinents pour évaluer quantitativement la qualité de la parole et comprendre l'état du larynx du locuteur. Divers indices acoustiques sont utilisés pour caractériser la parole produite par les personnes atteintes de dysphonie.

I.6.1 Les troubles de la parole et de la voix

Dans la communication parlée, phonétique et langage entretiennent une relation très étroite. La parole utilise la voix comme support acoustique pour exprimer le langage. La confusion des voix, de la parole et du langage est indiscernable dans certaines émotions, qui sont souvent étroitement liées. L'incapacité à produire un son clair et modulé interfère avec l'ordre des syllabes et donc la compréhension des mots. En conséquence, les troubles de la parole créent des barrières linguistiques et entraînent en même temps une diminution de la qualité de la parole. On distingue trois grands types de troubles liés à la voix, à la parole et au langage :

- La dysphonie est une qualité de parole anormale. Par exemple, l'enrouement peut résulter d'une paralysie, d'une voix tendue ou d'une laryngite.
- La dysarthrie est un trouble du langage associé à une paralysie ou à une incoordination des muscles impliqués dans les articulations. Elle est d'origine neurologique.
- L'aphasie est un trouble du langage causé par des dommages au cortex cérébral. Les patients ne comprennent plus le sens des mots ou ne peuvent pas s'exprimer.

I.6.2 Evaluation des troubles de la voix et de la parole

On peut distinguer deux approches pour l'évaluation des troubles de la parole et de la voix. La première méthode dite subjective est basée sur la perception auditive de la voix et de la parole. La deuxième méthode dite objective est basée sur l'acquisition de mesures physiques à l'aide de capteurs et l'application de techniques de traitement du signal sur les mesures extraites pour établir un bilan.

I.6.2.1 Evaluation subjective

On distingue deux formes d'évaluation subjective de la voix :

- celle du praticien : c'est l'analyse perceptive (ou perceptuelle).
- celle du patient : c'est l'auto-évaluation par questionnaire.

Pour l'analyse perceptive, le praticien doit juger la qualité de la voix et ses caractéristiques perceptives. Il peut pour cela utiliser différentes échelles dont la plus utilisée est l'échelle GRBASI développée par Hirano (1981) et complétée par la suite par Dejonckere (1996) qui évalue six paramètres :

- « G » (Grade global de dysphonie) pour caractériser l'impression générale de sévérité de la dysphonie.
- « R » (Raucité) pour caractériser la raucité qui réfère à l'impression d'irrégularité de vibration.
- « B » (Caractère soufflé) pour caractériser le souffle qui évalue la présence de fuite d'air altérant la voix, par manque d'accolement glottique phonatoire.
- « A » (Asthénie) pour l'asthénie qui traduit le manque de puissance, la faiblesse de la voix.
- « S » (Forçage) pour caractériser le serrage, l'impression d'effort et de tension, notamment lorsqu'il y a un forçage vocal.
- « I » (instabilité) pour mesurer l'instabilité qui correspond aux variations de la qualité de la voix d'un instant à l'autre.

Pour chaque paramètre, le praticien cote selon quatre degrés de sévérité, allant de 0 (normal / absence de dysphonie) à 3 (présence maximale de dysphonie). Ainsi, plus le score GRBASI est élevé, plus l'altération vocale est jugée importante par le professionnel [14].

L'évaluation perceptive est essentielle car l'oreille est l'outil le plus puissant et le plus accessible pour juger la qualité de la voix, et souvent cette analyse permettra d'évaluer l'évolution après une thérapie. Cependant, elle n'est pas suffisante car elle est biaisée. La variabilité significative du score inter-évaluateur (jugement de la même voix par deux évaluateurs) et même intra-évaluateur (jugement du même évaluateur à des moments différents) a été réduite par l'utilisation de comités d'audition, mais ceux-ci sont difficiles à mettre en œuvre dans la pratique. La complémentarité des évaluations instrumentales est donc essentielle, et les évaluations dites « objectives » compensent cette importante subjectivité.

I.6.2.2 Evaluation objective

L'évaluation objective utilise les techniques de traitement automatique de la parole (TAP). L'objectif est d'extraire des indices acoustiques pertinents pour caractériser la voix, fournissant ainsi des informations sur l'état du larynx du locuteur. Les cliniciens préfèrent les signaux acoustiques associés aux caractéristiques perceptives. Cependant, il est vain de rechercher une cohérence perceptive-acoustique lorsque les évaluations perceptives ne sont pas suffisamment fiables. La recherche dans la caractérisation des troubles de la parole et de la voix a pour but le développement de méthodes analytiques spécifiques pour estimer les perturbations dans les signaux de parole. Il existe plusieurs façons d'évaluer objectivement les troubles de la voix. Les méthodes analytiques présentées dans cette section se répartissent en deux catégories, selon qu'elles utilisent des signaux vocaux ou des estimations de sources glottiques pour calculer des indices acoustiques.

Les indices acoustiques utilisés lors de l'analyse représentent les caractéristiques temporelles et spectrales restantes. Ces indices sont :

I.6.2.2.1 Indices de perturbation de la fréquence fondamentale

- **Jitter**: C'est la mesure des variations d'un cycle à l'autre du signal d'excitation glottique. Le jitter est affectée principalement par le manque de contrôle de la vibration des cordes. La voix des locuteurs pathologiques est souvent caractérisée par des valeurs de jitter plus élevées.

La perturbation du jitter peut être donnée par cinq paramètres:

- **le jitter absolue (jitta)** : C'est la perturbation absolue calculée comme la différence absolue moyenne entre des périodes consécutives, en secondes.

- **le jitter locale ou relative (jitt)** : C'est la différence absolue moyenne entre des périodes consécutives, divisée par la période moyenne, en pourcent.
- **la perturbation moyenne relative (rap)** (the Relative Average Perturbation) : C'est la différence absolue moyenne entre une période et la moyenne de celle-ci et de ses deux voisins, divisée par la période moyenne, en pourcent.
- **La perturbation de période à cinq points (ppq5)** : (five-point Period Perturbation Quotient) : C'est la différence absolue moyenne entre une période et la moyenne de celui-ci et de ses quatre voisins les plus proches, divisée par la période moyenne, en pourcent.

I.6.2.2.2 Indices de perturbations d'amplitude

- **Shimmer**: C'est une variation des amplitudes de périodes consécutives, qui peut également être mesurée en soustrayant l'amplitude de la séquence de périodes de pitch à son voisin ou à des combinaisons de ses voisins. On distingue aussi quatre mesures liées au shimmer :
 - **Le shimmer absolu ou local (ShdB)** : C'est le logarithme absolu moyen en base 10 de la différence entre les amplitudes de périodes consécutives, multipliées par 20 et exprimées en décibels (dB).
 - **Le shimmer local (Shim)** : C'est la différence absolue moyenne entre les amplitudes de périodes consécutives, divisée par l'amplitude moyenne, exprimée en pourcent.
 - **L'Amplitude à trois points Quotient de perturbation (apq3)** : C'est la différence absolue moyenne entre l'amplitude d'une période et la moyenne des amplitudes de ses voisines, divisée par l'amplitude moyenne, exprimée également en pourcent.
 - **l'amplitude à cinq points Quotient de perturbation (apq5)** : C'est la différence absolue moyenne entre l'amplitude d'une période et la moyenne des amplitudes de celle-ci et de ses quatre plus proches voisins, divisée par l'amplitude moyenne, exprimée également en pourcent.

I.6.2.2.3 Rapport harmonique sur bruit (HNR)

Le rapport harmonique sur bruit (Harmonics-to-Noise Ratio : HNR) mesure le rapport entre l'énergie harmonique qui résulte des interruptions quasi-périodiques du

passage de l'air entre les cordes vocales, et l'énergie du bruit qui correspond aux inter-harmoniques, c'est-à-dire le passage turbulent et continu de l'air à travers la glotte [15].

$$HNR = 10 \log_{10} \left(\frac{AC_v(T)}{AC_v(0) - AC_v(T)} \right) \quad (I.8)$$

où : $AC_v(0)$ est le coefficient d'autocorrélation à l'origine qui est aussi l'énergie du signal, et $AC_v(T)$ est la composante d'autocorrélation correspondante à la période du pitch. La différence $AC_v(0) - AC_v(T)$ peut-être supposée comme l'énergie du bruit.

I.6.2.2.4 Rapport signal sur bruit (SNR)

Le rapport signal sur bruit (Signal-to-Noise Ratio : SNR) est un indicateur de la qualité de la transmission du signal. Si ce rapport est positif, la parole est d'un degré supérieur au bruit. Par contre, si ce rapport est négatif, le bruit est plus intense que la parole. Il peut s'exprimer par le rapport des puissances du signal (P_s) et du bruit (P_n), ou par le rapport des valeurs RMS du signal (S) et du bruit (N) [16].

$$SNR = 10 \log \left(\frac{\sum_{n=1}^N f(n)^2}{\sum_{n=1}^N [f(n) - \hat{f}(n)]^2} \right) \quad (I.9)$$

où : $f(n)$ est un signal contenant du bruit

$\hat{f}(n)$ est le signal débruité

N est la longueur du signal

I.6.2.2.5 Rapport signal à dyspériodicité

Rapport global signal/dyspériodicité dans (Qi et al. 1999), le SDR, exprimé en dB, a été défini comme le logarithme du rapport de l'énergie du signal et de l'énergie de la dyspériodicité, moins 1. Alors, le logarithme n'est pas défini lorsque l'énergie de la dyspériodicité est supérieure à l'énergie du signal. Ici, le SDR est donc calculé comme suit [17] :

$$SDR = 10 \log \left[\frac{\sum_{n=0}^{L-1} x^2(n)}{\sum_{n=0}^{L-1} e^2(n)} \right] \quad (I.10)$$

où : $e(n)$ La dyspériodicité

I.7 Conclusion

Dans ce chapitre, nous avons présenté les notions de base du signal de parole. Les indices acoustiques présentent un intérêt particulier dans le cadre de l'analyse de la parole

pour l'évaluation clinique de la qualité de la voix. En effet, dans l'évaluation clinique de la voix, les méthodes d'analyse acoustiques basées sur des mesures objectives sont utilisées en parallèle aux méthodes d'évaluation auditive pour la documentation et l'évaluation des troubles de la voix. Grâce à la relation entre la fonction du larynx et la production de la voix, l'analyse du signal acoustique permet de détecter et d'évaluer la sévérité de plusieurs pathologies laryngées. Étant donné que dans un système de télémédecine, les signaux sont compressés avant transmission, le processus de compression peut affecter la performance de l'analyse acoustique à distance.

chapitre II :

Systemes de télémédecine

II.1 Introduction

Grâce au progrès de la technologie et de la médecine, les humains ont tendance à vivre plus longtemps. Cela a accru la pression sur les systèmes de santé du monde entier pour qu'ils fournissent des soins de santé de meilleure qualité à un plus grand nombre de patients. Une plus grande demande de services de soins de santé a incité les chercheurs à rechercher de nouvelles façons d'organiser et de fournir les services de soins de santé. La télémédecine, en tant que nouveau domaine de recherche, permet l'utilisation de systèmes multimédias comme moyen afin d'augmenter la disponibilité des soins pour les patients en plus des stratégies d'économie de temps et d'argent. La télémédecine permet aux patients d'être examinés et traités, tandis que le fournisseur de soins de santé et le patient se trouvent à des emplacements physiques différents. Grâce aux technologies de télémédecine, les futurs hôpitaux fourniront des services de soins de santé aux patients du monde entier à l'aide de systèmes multimédias et de signaux pouvant être acquis à distance. La transmission, le stockage et le traitement des signaux et des images sont les composantes majeures de la télémédecine [18].

Dans ce chapitre, on donne une définition de la télémédecine ainsi que la configuration de base d'un système de télémédecine, ces différents actes, ces avantages et ces inconvénients. On présente aussi l'application de la télémédecine pour l'analyse et l'évaluation des troubles de la voix.

II.2 Définition de la télémédecine

La télémédecine est un terme générique désignant toutes les formes d'échange d'informations médicales [19], on peut dire que la télémédecine est une forme de pratique médicale à distance utilisant les technologies de l'information et de la communication pour mettre en relation des patients ou des professionnels de la santé (médecin, infirmière, chirurgien-dentiste...). Elle peut renforcer la coopération internationale, combiner les pratiques médicales et faciliter l'intégration du cyber santé dans les systèmes de santé. Les développements technologiques récents (lignes téléphoniques fixes et mobiles, Internet, satellite, etc.) permettent de dispenser des soins de santé à distance et d'échanger des informations médicales connexes. La télémédecine a aujourd'hui trouvé de nombreux champs d'applications, et se décline en différents termes dont il est difficile de déterminer une typologie unanime [20,21]. La Figure II.1 montre la configuration d'un système de télécardiologie.

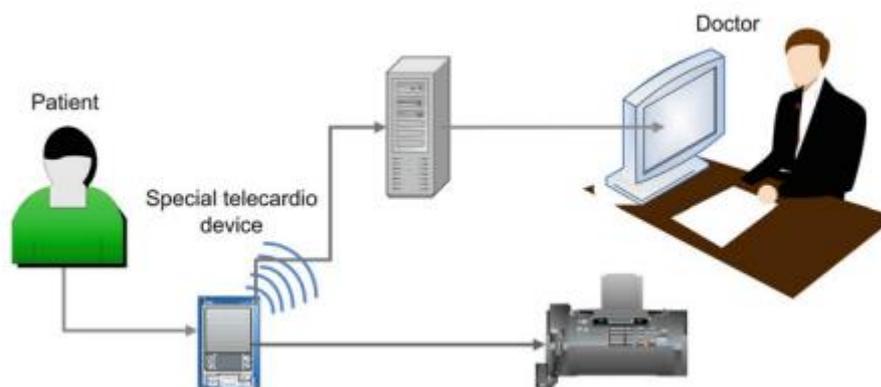


Figure II.1 : Exemple de systèmes multimédias en télécadiologie [18].

II.3 Les actes de la télémédecine

Pour clarifier les responsabilités médicales engagées, il importe de simplifier la définition des actes médicaux de la télémédecine en prenant en compte les définitions adoptées par l'ensemble des pays, notamment européens. Le champ de la télémédecine (et non de la télésanté) peut être couvert par les actes principaux que sont la téléconsultation, la télé expertise et la télésurveillance comme montré sur la Figure II.2. La téléassistance est un acte qui n'est pas toujours médical [22].

II.3.1 La télé expertise

La télé expertise a été limitée souvent dans sa définition aux échanges entre spécialistes pour obtenir un deuxième avis. Il semble souhaitable, par souci de simplification, d'élargir cette définition à tout acte diagnostic et/ou thérapeutique qui se réalise en dehors de la présence du patient. L'acte médical de télé expertise se décrit comme un échange entre deux ou plusieurs médecins qui arrêtent ensemble un diagnostic [22].

II.3.2 La télé consultation

La télé consultation, a pour objet de permettre à un professionnel médical de donner une consultation à distance à un patient. Un professionnel de santé peut être présent auprès du patient et, le cas échéant, assister le professionnel médical au cours de la téléconsultation. Les psychologues peuvent également être présents auprès du patient [23].

II.3.3 Téléassistance

La téléassistance peut-être un acte médical lorsqu'un médecin assiste, à distance, un autre médecin en train de réaliser un acte médical ou chirurgical, voire dans le cadre de

l'urgence [22], aide un secouriste ou toute personne assistante à une personne en danger en attendant l'arrivée d'un médecin [24].

II.3.4 La télésurveillance

La télésurveillance utilise des techniques de transfert et des thérapies pour recueillir des données cliniques, radiologiques ou biologiques recueillies par le patient lui-même ou par un professionnel de la santé dans un endroit éloigné [25] [23]. La surveillance médicale a pour but de permettre à un professionnel de la santé d'interpréter les données nécessaires au suivi médical à distance d'un patient et de veiller à ce qu'elles soient échangées en toute sécurité à distance [26].

II.3.5 Téléformation

L'utilisation de l'outil informatique en particulier pour l'aide à la formation continue des médecins : contacts professionnels via le réseau, consultation des informations médicales (banque de données, imagerie, suivi d'études épidémiologiques et d'essais cliniques), consultation de cours de formation et visioconférences dans les universités (téléenseignement) et réunions. En particulier, l'application dénommée couramment télésurveillance médicale à domicile est fondamentale pour l'amélioration de la qualité de soins et de vie des personnes nécessitant des soins ou une attention particulière. Elle vise à mettre en place dans l'habitat d'une personne un dispositif qui permet de capturer des informations sur son état de santé, afin de rendre possible pour le praticien un diagnostic, voire une aide au patient à distance [27].



Figure II.2 : Les types d'application de la télémédecine.

II.4 La chaîne de télémédecine

Le système de la transmission numérique des données en médecine comprend tout une chaîne :

- **Le patient:** Le patient est la source de destination d'information médicale, le D.T.E (Data Terminal Equipment) est chargé de prélever sur le corps humain l'information médicale.
- **Codeur/Décodeur:** Permet la transition de l'information médicale vers le pc locale.
- **Pc Local:** Permet de présenter l'information médicale au praticien de la médecine et de stocker ces informations dans un système d'archivage et d'envoyer une plate de forme de traitement numérique et le transfert de l'information via un protocole de communication.
- **D.C.E:** (Data Communication Equipment) : Chargé d'adapter le signal informationnel au canal de transmission et de transférer les données médicales vers les terminaux distants (Pc Distant) via le canal de transmission au moyen des techniques hauts débits à titre d'exemple le réseau ADSL.

L'objectif de telles plateformes de services de télémédecine est de permettre aux patients de vivre dans des conditions plus performantes, dans un environnement de confort et de sécurité. Ainsi, cette plateforme envisagée permet, à tout moment et en tout lieu, à un patient d'être en contact permanent avec son praticien traitant. En effet, ils pourraient bénéficier de la sécurité d'un suivi médical 24h/24h, sans l'inconvénient de la prise en charge hospitalière et sans dépenses excessives. Il s'agit de détecter et de prévenir l'occurrence de situations critiques ou la dégradation de l'état de santé d'une personne. Le patient n'est alors plus contraint de renoncer à distance (domicile...) et à la vie en société. Il conserve une large autonomie dans son environnement social et privé, tout en bénéficiant de services préventifs de santé [28].

II.5 Avantages et inconvénients de la télémédecine

II.5.1 Avantages

Les avantages de la télémédecine sont nombreux :

- Une consultation en télémédecine peut se faire par téléphone ou par appel vidéo.
- La télémédecine peut aussi prendre une place plus importante lorsque le système de santé est saturé.
- La réduction du temps passé en salle d'attente.
- Une réponse aux difficultés de mobilité.
- Moins de risques de contamination.

II.5.2 Inconvénients

- La télémédecine possède bien sûr certaines limites, comme dans des cas où l'examen physique d'un patient est nécessaire.
- Mauvaise connexion Internet.
- Risque de fuite des données de santé personnelles.
- Il y'a une catégorie des personnes qui ne sont pas équipés pour cette pratique.

II.6 Le service de télémédecine pour les troubles de la parole et de la voix

II.6.1 Configuration du système

L'analyse du signal de parole offre un moyen privilégié pour l'évaluation clinique de la qualité de la voix en vue d'un diagnostic et d'une documentation quantitative des

pathologies du larynx. Une analyse acoustique présente plusieurs avantages. En plus du coût peu élevé du système d'évaluation, elle est simple à mettre en œuvre, présente un caractère non invasif et documente quantitativement le degré d'enrouement perçu par le clinicien.

L'objectif de l'analyse du signal acoustique est d'extraire des indices pertinents permettant de déterminer les caractéristiques de la voix afin de renseigner sur l'état du larynx du locuteur. Dans ce contexte, plusieurs indices acoustiques sont utilisés pour caractériser la parole produite par des locuteurs dysphoniques.

La télémédecine a le potentiel de fournir des services à domicile ou dans la communauté locale par vidéoconférence et par le biais d'activités de thérapie interactives sur ordinateur. Cette forme de prestation de services a la capacité d'optimiser les résultats fonctionnels en facilitant la généralisation des effets du traitement dans l'environnement quotidien de la personne et de permettre un suivi à long terme des comportements de communication et de déglutition. Un certain nombre d'applications de télémédecine basées sur l'image ont été utilisées dans la prise en charge des troubles neurogènes de la parole et du langage chez l'adulte, du bégaiement, des troubles de la voix, des troubles de la parole et du langage chez l'enfant, de la laryngectomie et des troubles de la déglutition. La poursuite du développement de ces applications et d'autres thérapies informatisées, des analyses coûts-avantages et coût-efficacité et une formation professionnelle sont nécessaires si la télémédecine doit devenir une partie intégrante de la pratique de l'orthophonie [29].

La configuration d'un système de télémédecine pour la détection et l'évaluation des troubles de la parole et de la voix est illustrée par la Figure II.3.

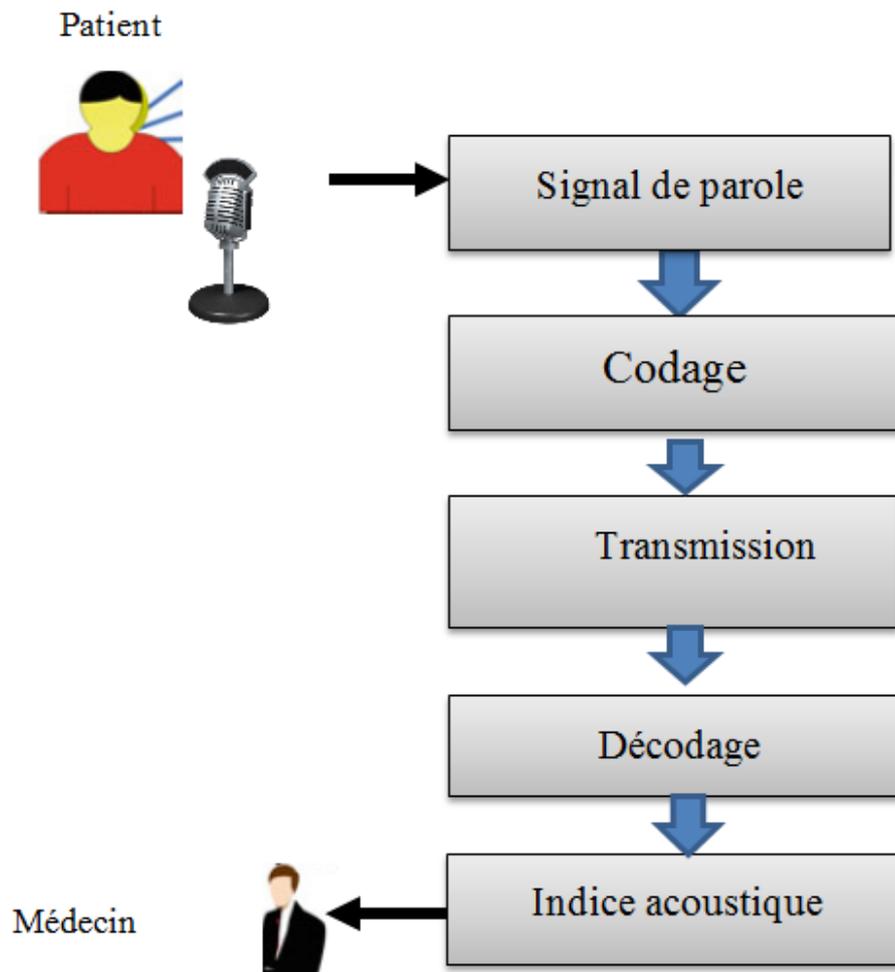


Figure II.3 : Chaîne de transmission de signal de parole pour les services de télémédecine.

II.6.2 Codage et compression de la parole

Le traitement de signal de la parole nécessite toujours en premier lieu une conversion de signal vocal en signal électrique par le microphone. La numérisation du signal permet de construire une représentation discrète d'un objet du monde réel. Ce procédé implique d'abord un échantillonnage suivi d'une opération de quantification puis en dernière étape le codage.

Le codage ou la compression permet de réduire la taille des données à envoyer, à traiter ou à stocker en supprimant les redondances contenues dans le message afin de réduire le débit et rendre la transmission plus rapide mais sans une perte d'informations initiales. Il existe trois critères sur lesquels se base les algorithmes de compression [30]:

- Le taux de compression : c'est le rapport de la taille du fichier compressé sur la taille du fichier initial. Il est généralement exprimé en pourcentage. Le taux de compression est calculé par :

$$\tau = 1 - (b/a) \quad (\text{II.1})$$

où b est la taille du fichier compressé et a la taille du fichier initial.

- La qualité de compression : sans ou avec pertes (avec le pourcentage de perte).
- La vitesse de compression et de décompression.

II.6.3 Classification des algorithmes de compression

Le critère de classification le plus pertinent est basé sur la perte des données. On peut distinguer deux types de compression qui seront présentés comme suit [31] :

II.6.3.1 Compression sans perte

Ce type de compression est très utile, car certains types de fichiers, notamment les fichiers de données, ne peuvent pas se permettre de ne perdre aucun bit de donnée pour ne pas perdre l'intégralité de leur sens. Les algorithmes de compression sans perte (connu aussi sous le nom de non destructible, réversible, ou conservative) sont des techniques permettant une reconstitution exacte de l'information après le cycle de compression / décompression [32].

II.6.3.2 Compression avec perte

L'algorithme de compression repère donc les sons "dominants" et retire toutes les données relatives aux sons "dominés" pour ne transmettre que ce qui est perceptible, il élimine l'information redondante et introduit une dégradation indiscernable à l'oreille avec un taux de compression très élevé. Donc, elle ne s'applique qu'aux données « perceptibles », les données originales ne peuvent pas être retrouvées, donc la perte d'information est irréversible c.-à-d. non conservative. Les données cachent les bruits de certaines fréquences et les différences de phase ne sont pas détectées par l'oreille humaine pour la plupart des fréquences. Tout cela donne la possibilité de comprimer la parole pour obtenir des résultats de bonne qualité avec un stockage minimal [31]. Exemple : la technologie de compression MP3 (ou MPEG Audio Layer 3).

II.6.4 La technologie de compression MP3

II.6.4.1 Principe de fonctionnement

La technologie MP3 (ou MPEG Audio Layer 3) est un algorithme de compression audio qui permet de réduire la taille d'un fichier audio. Cette technologie a été développée en 1992 par le Moving Picture Experts Group (MPEG), groupe d'experts chargé du développement de normes internationales pour la compression, la décompression, le traitement et le codage de la vidéo et de l'audio.

La technologie MP3 n'est en réalité pas une compression mais plutôt une suppression d'informations dans l'objectif de simplifier et alléger un fichier. Cette technologie repose sur deux aspects centraux : la capacité auditive humaine d'une part et un modèle psychoacoustique d'autre part.

- **Suppression des fréquences inaudibles**

Tout d'abord, il faut savoir que l'oreille humaine a une capacité limitée : elle perçoit uniquement les sons compris dans la fourchette de fréquences 20Hz - 20KHz. La technologie MP3 consiste alors à supprimer les données correspondant aux fréquences inaudibles. Ainsi, le fichier d'origine est allégé, débarrassé des données imperceptibles, sans que l'utilisateur distingue la modification.

- **Suppression des fréquences masquées**

La compression au format MP3 utilise d'autre part un modèle psychoacoustique de l'effet de « masque » : si deux fréquences d'intensités différentes sont présentes en même temps, la plus faible peut être moins perçue que l'autre, elle sera donc « masquée » par la fréquence plus puissante. Pour utiliser une comparaison plus marquante, imaginez que vous êtes dans votre jardin et que vous entendez des oiseaux. Si un avion passe au-dessus de votre tête (même très haut), vous n'entendrez plus les oiseaux car le son qu'ils produisent est beaucoup plus faible que celui généré par l'avion. C'est comme si les oiseaux n'existaient plus ou avaient arrêté de chanter. On voit bien qu'il n'est pas nécessaire de coder toutes les fréquences présentes dans un morceau pour que l'oreille humaine le perçoive correctement quand même. La technologie MP3 consiste alors à effacer les données qui correspondent à ces fréquences masquées que l'oreille humaine ne discerne pas. La taille du fichier est donc encore réduite sans que l'utilisateur le distingue.

II.6.4.2 Algorithmes de codage

Le MP3 utilise également l'algorithme de Huffman (1952) comme méthode de codage des informations. Cette méthode est utilisée dans tous les algorithmes de compression (compression de fichiers texte, compression d'images, compression de sons).

Elle repose sur l'utilisation d'un code de longueur variable et la probabilité d'apparition d'un événement (en l'occurrence ici d'une fréquence). Plus une fréquence apparaît souvent, plus son code sera court (nombre de bits faible pour la représenter). Le fichier est lu une première fois et on dresse un tableau des fréquences apparaissant et le nombre de fois où elles apparaissent. On en déduit le code approprié. Ce codage est utilisé en dernier lieu. C'est la phase finale de la compression. C'est un codage non destructeur.

L'utilisation de toutes les possibilités de réduction évoquées dépend de la place que l'on souhaite accorder à une minute de son compressé et donc du taux de compression que l'on va devoir appliquer. Pour l'encodage de fichiers sont en MP3, on parle plutôt en terme de taux de bits (bit rate) que de taux de compression. Le taux de bits est le nombre de bits autorisés en 1 seconde. On aura donc la relation suivante : plus on veut compresser un morceau (pour qu'il prenne le moins de place possible) plus le taux de bits à choisir devra être faible [33].

II.7 Conclusion

La télémédecine est un outil d'amélioration de la qualité des soins, elle peut rendre de différents services aux malades. Dans ce chapitre on a représenté la configuration d'un système de télémédecine, son utilisation pour la détection et l'évaluation des troubles de la parole et de la voix. La compression du signal de parole est l'un des traitements essentiels lors de la transmission mais elle peut réduire la qualité du signal lors de la reconstruction.

chapitre III :

Acquisition comprimée

III.1 Introduction

L'acquisition comprimée est un nouveau domaine de recherche, qui a été introduit en 2006, et depuis lors est déjà devenu un concept clé dans divers domaines des mathématiques appliquées, de l'informatique et du génie électrique. Il prédit de manière surprenante que les signaux de grande dimension, qui permettent une représentation parcimonieuse dans une base appropriée peuvent être récupérés à partir de ce qui était auparavant considéré comme des mesures linéaires très incomplètes en utilisant des algorithmes efficaces. Dans ce chapitre, on introduit la théorie de base de l'acquisition comprimée et on donne un exemple d'illustration de la compression et de la reconstruction du signal.

III.2 Principe de l'acquisition comprimée

L'acquisition comprimée (CS : compressive sensing) est un nouveau concept type de la théorie d'échantillonnage qui stipule que des signaux et des images parcimonieux peuvent être reconstruits à partir de ce qui était auparavant considérée comme une information incomplète. Des algorithmes efficaces tels que la minimisation l_1 peuvent être utilisés pour la récupération. La théorie a de nombreuses applications potentielles dans le traitement du signal et de l'imagerie. En fait, l'idée est d'acquérir seulement quelques mesures permettant une reconstruction (sous certaines contraintes) plutôt que d'acquérir un grand nombre d'échantillons au taux de Nyquist, puis d'effectuer une compression comme montre Figure III.1. Cela peut réduire les ressources nécessaires pour échantillonner des signaux dont la plupart de leurs entrées sont de toute façon nulles.

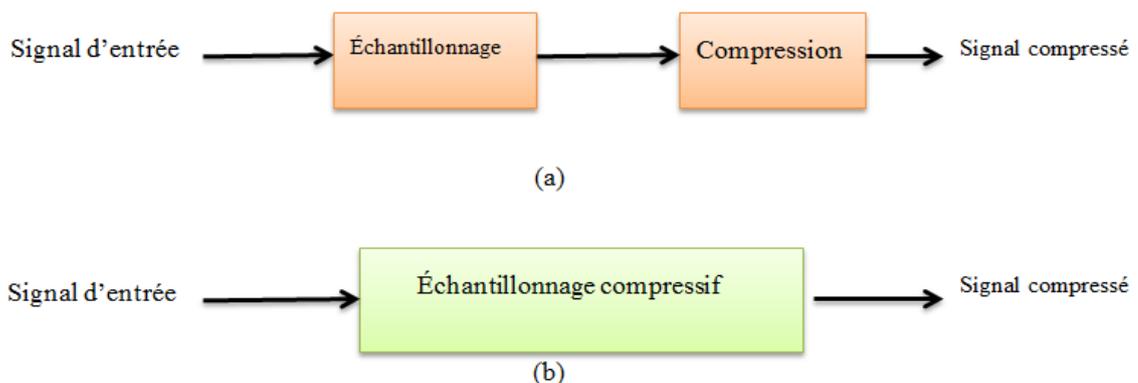


Figure III.1 : (a) Échantillonnage classique puis compression et (b) acquisition comprimée.

Le contexte théorique de l'acquisition comprimée a des liens avec de nombreux autres domaines tels que l'algèbre linéaire numérique, la théorie d'optimisation de l'optimisation, la théorie des matrices aléatoires, l'analyse harmonique appliquée, etc. Le processus de l'acquisition comprimée est composé de trois parties principales telles que présentées dans Figure III.2 :

1. Représentation parcimonieuse dans une base appropriée pour obtenir un vecteur avec seulement quelques valeurs non nulles.
2. Acquisition à l'aide d'une matrice de mesure pour obtenir les mesures compressées.
3. Reconstruction du signal comprimé à l'aide d'un algorithme de reconstruction pour récupérer le vecteur parcimonieux.

Le signal s de longueur N est parcimonieux en le projetant sur la base de représentation Ψ qui peut être une base de Fourier, d'ondelettes, etc. On obtient ainsi le vecteur parcimonieux θ (de longueur N) qui ne contient que K valeurs non nulles, on dit donc qu'il est K -parcimonieux.

L'acquisition est réalisée en multipliant la matrice de mesure Φ de longueur $M \times N$ par le signal s . On obtient donc le signal acquis y (vecteur de mesure) qui ne contient que M mesures, tel que $K < M \ll N$ [34].

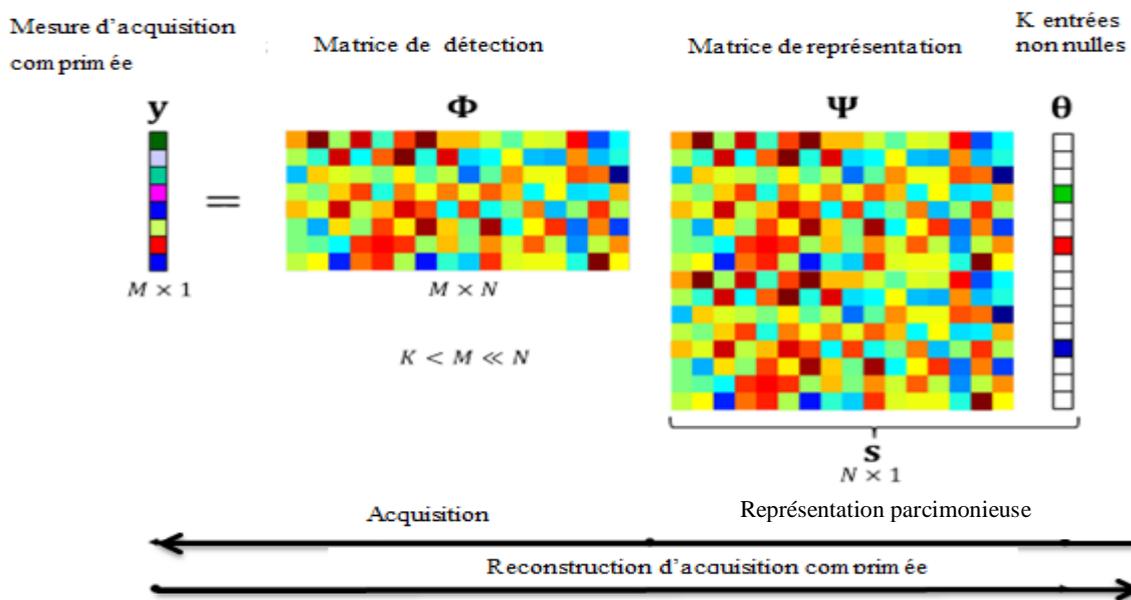


Figure III.2 : Principe de l'acquisition comprimée.

III.3 Parcimonie

De nombreux signaux naturels ont des représentations concises lorsqu'ils sont exprimés dans une base appropriée. Considérons, par exemple, l'image de la Figure III.3 (a) et sa transformée en ondelettes dans la Figure III.3 (b). Bien que presque tous les pixels de l'image aient des valeurs non nulles, les coefficients d'ondelettes offrent un résumé concis : la plupart des coefficients sont petits et les coefficients relativement peu importants capturent la plupart des informations.

Mathématiquement parlant, on a un vecteur $f \in R^n$ (tel que l'image à n-pixels de la Figure III.2 que nous développons dans une base orthonormée (telle qu'une base d'ondelettes) $\Psi = [\Psi_1 \Psi_2 \dots \dots \Psi_n]$ comme suit :

$$f(t) = \sum_{i=1}^n x_i \Psi_i(t) \quad (\text{III.1})$$

où \mathbf{x} est la suite de coefficients de f , $x_i = \langle f, \Psi_i \rangle$. Ce sera pratique pour exprimer f comme $\Psi \mathbf{x}$ (où Ψ est la matrice $n \times n$ ayant $\Psi_1, \dots, \dots, \Psi_n$ comme colonnes).

L'implication de la parcimonie est maintenant claire: Lorsqu'un signal a une représentation parcimonieuse, on peut écarter les petits coefficients sans grande perte de perception. Formellement, considérons $f_s(t)$ obtenu en ne gardant que les termes correspondant aux S plus grandes valeurs de (x_i) . Par définition, $f_s := \psi x_s$, où, x_s est le vecteur de coefficients (x_i) avec tous mis à zéro sauf les S plus grandes valeurs. Ce vecteur est parcimonieux au sens strict puisque toutes ses entrées sauf quelques-unes sont nulles ; nous appellerons S -parcimonieux de tels objets avec au plus S entrées non nulles. Puisque ψ est une base orthonormée, on a $\|f - f_s\|_{l_2} = \|x - x_s\|_{l_2}$ et si \mathbf{x} est parcimonieux ou compressible dans le sens où les valeurs triées des (x_i) décroissent rapidement, alors \mathbf{x} est bien approximé par x_s et, par conséquent, l'erreur $\|f - f_s\|_{l_2}$ est petite. En clair, on peut négliger une grande partie des coefficients sans grande perte. La Figure III.3 (c) montre un exemple d'une image mégapixel où la perte de perception est à peine perceptible à son approximation obtenue en négligeant 97,5% des coefficients. Ce principe est, bien sûr, ce qui sous-tend la plupart des codeurs avec perte tels que JPEG-2000 [35] et de nombreux d'autres [36].

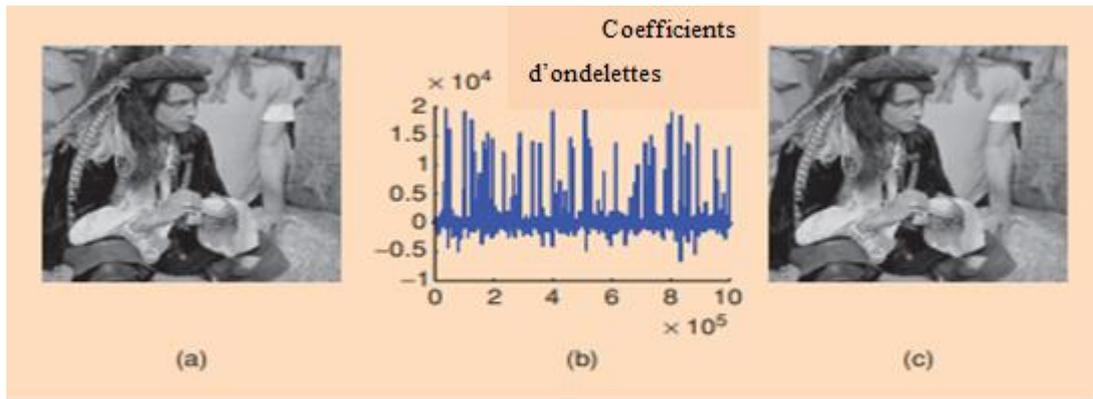


Figure III.3 : (a) Image mégapixel originale avec des valeurs de pixel dans la plage $[0,255]$ et (b) ses coefficients de transformée en ondelettes (arrangés dans un ordre aléatoire pour une meilleure visibilité). (c) La reconstruction obtenue en mettant à zéro tous les coefficients de l'expansion en ondelettes.

III.4 Acquisition du signal

L'acquisition comprimée du signal s correspond à la collecte d'un petit nombre de mesures linéaires de s . Cela peut être représenté comme :

$$y_i = \langle s, \Phi_i \rangle \quad i = 1, \dots, M \quad (\text{III.2})$$

où Φ_i est la i ème ligne de la matrice de mesure Φ de dimensions $M \times N$ avec

$K < M \ll N$, et M est le nombre de mesures nécessaires à la reconstruction. $\langle \cdot, \cdot \rangle$ désigne le produit scalaire. Le signal y , appelé vecteur d'observation (ou vecteur de mesure) peut aussi s'écrire (sous forme matricielle) sous la forme :

$$y = \Phi_s = \Phi\Psi\theta = A\theta \quad (\text{III.3})$$

où θ est le vecteur des coefficients creux du signal. On dit que y contient des mesures de compression de s .

Une reconstruction parfaite nécessite que Φ et Ψ obéissent à deux conditions appelées (a) condition d'incohérence et (b) propriété d'isométrie restreinte.

III.4.1 Incohérence

Soit (Φ, Ψ) un couple de bases de \mathbf{R}^N . La première base Φ (également connue sous le nom de matrice de mesure) est utilisée pour prendre une mesure du signal s , et la seconde base Ψ (également connue sous le nom de matrice de parcimonie ou de représentation) est utilisée pour représenter de manière parcimonieuse s .

L'incohérence signifie que la cohérence mutuelle entre la matrice de mesure Φ et la base creuse Ψ est faible. La cohérence mutuelle mesure la plus grande corrélation entre deux éléments quelconques de Φ et Ψ . En d'autres termes, c'est la valeur absolue maximale des inters-corrélations entre les colonnes de Φ et Ψ . Elle peut être définie comme :

$$\mu(\Phi, \Psi) = \alpha \max_{1 \leq k, j \leq N} |\langle \Phi_k, \Psi_j \rangle| \quad (\text{III.4})$$

où Φ_k et Ψ_j sont respectivement les $k^{\text{ième}}$ et $j^{\text{ième}}$ colonnes de Φ et Ψ . Selon divers auteurs, α peut prendre les valeurs N , \sqrt{N} ou une valeur normalisée à un. La valeur de α dépend des propriétés algébriques des matrices choisies Φ et Ψ .

La cohérence sert de caractérisation grossière du degré de similarité (ou «corrélation») entre la parcimonie et les systèmes de mesure. La valeur $\mu = 1$ est le minimum (valeur la plus basse) de la cohérence, ce qui signifie que les lignes de la matrice de mesure Φ doivent être étalées dans le domaine Ψ . En d'autres termes, plus la cohérence est faible, plus l'incohérence est grande. Cette idée apparaît dans l'acquisition comprimée lors du choix de la matrice de mesure : la matrice de mesure doit être incohérente avec la base dans laquelle la parcimonie est exprimée. Pour plus de précision, considérons deux exemples particuliers de matrices Φ et Ψ :

- Soit Φ la matrice de mesure correspondante à un échantillonnage classique dans le temps ou dans l'espace, et Ψ la base de Fourier. Le couple temps-fréquence satisfait $\mu(\Phi, \Psi) = 1$, et par conséquent, il y a une incohérence maximale. La notion d'incohérence prolonge la dualité entre temps et fréquence : un signal très localisé dans le temps est étalé en fréquence (principe d'incertitude).
- La matrice de mesure Φ est choisie comme une matrice aléatoire gaussienne dont les éléments sont indépendants les uns des autres et la matrice de représentation Ψ est une base d'ondelettes orthogonales (par exemple ondelettes de Daubechies). Ce choix est motivé par le fait que si la matrice de mesure est aléatoire, alors elle est largement incohérente avec toute base fixe Ψ [34].

Propriété d'isométrie restreinte

C'est un problème clé dans CS lors de l'étude de sa robustesse générale: la propriété dite d'isométrie restreinte (RIP : Restricted Isometry Property). Cela nécessite que tous les sous-ensembles de K colonnes tirés de la matrice produit A (définie comme $A = \Phi \Psi$)

soient en fait presque orthogonaux (ils ne peuvent pas être exactement orthogonaux puisque nous avons plus de colonnes que de lignes).

La propriété d'isométrie restreinte s'exprime comme suit :

$$1 - \delta_k \leq \frac{\|\Phi \Psi \theta\|_2^2}{\|\theta\|_2^2} \leq 1 + \delta_k \quad (\text{III.5})$$

où δ_k est la constante RIP d'ordre K , et $\|\cdot\|_2$ désigne la norme l_2 . On dit que la matrice A satisfait la propriété d'isométrie restreinte K avec une constante d'isométrie restreinte δ_k .

La propriété d'isométrie restreinte est la condition suffisante pour les matrices d'acquisition qui garantit les performances de la reconstruction du signal en termes d'efficacité et de robustesse. Comme les matrices aléatoires avec des entrées indépendantes et identiquement distribuées remplissent la condition RIP avec une probabilité élevée, ce sont les matrices de mesure les plus utilisées dans l'acquisition comprimée.

Depuis qu'il a été démontré que la matrice aléatoire gaussienne remplit les deux conditions lorsqu'elle est utilisée avec toute autre base, elle est considérée comme une matrice de mesure universelle qui peut être utilisée avec n'importe quelle base (y compris celles formées à l'aide de l'apprentissage par dictionnaire); d'où sa large utilisation [34].

III.4.2 Mauvaise matrice de mesure

Jusqu'à présent, nous avons décrit comment prendre de bonnes mesures compressées. La Figure III.4 montre un choix particulièrement mauvais de la matrice de mesure Φ correspondante aux M dernières colonnes de la base parcimonieuse Ψ . Dans ce cas, le produit $A = \Phi \Psi$ est une matrice identité $M \times M$ complétée par des zéros à gauche. Dans ce cas, tout signal S qui n'est pas actif dans les M dernières colonnes de Ψ est dans l'espace nul de A , et est totalement invisible aux mesures y . Dans ce cas, ces mesures entraînent des pertes d'informations importantes pour de nombreux vecteurs parcimonieux [37].

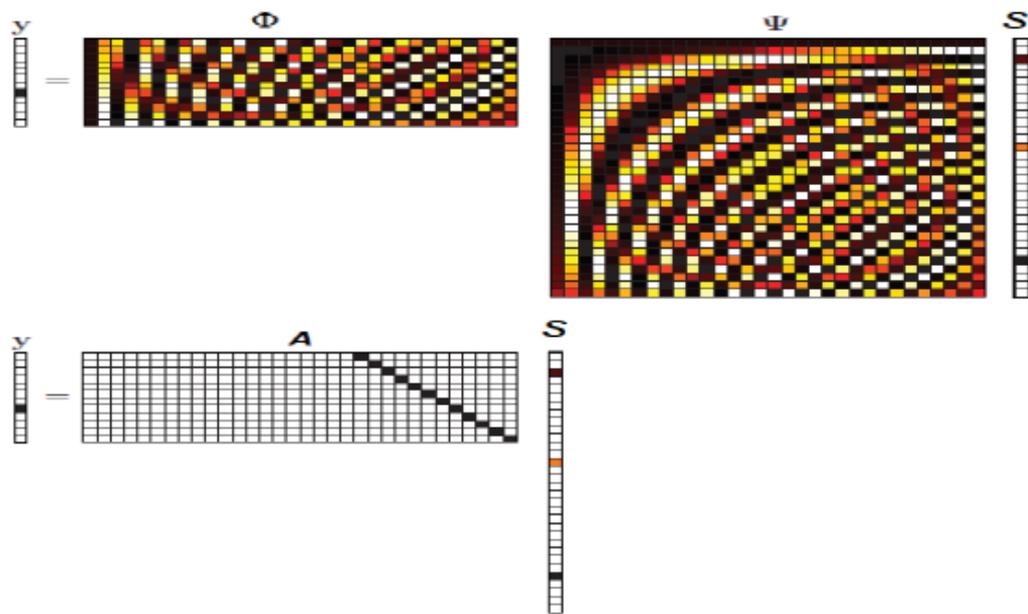


Figure III.4 : Exemple d'une mauvaise matrice de mesure.

III.5 Algorithmes de récupération : théorie de l'optimisation

Soit x un vecteur creux. Il est assez intuitif de récupérer x à partir de la connaissance de y en résolvant :

$$(P0) \min_x \|x\|_0 \text{ tel que } y = Ax \quad (\text{III.6})$$

Du fait de l'inévitable recherche combinatoire, cet algorithme est cependant NP-difficile [39]. L'idée principale de Chen, Donoho et Saunders dans l'article fondamental [38] était de substituer la norme « l_0 » par la norme convexe la plus proche, qui est la norme « l_1 ». Cela conduit au problème de minimisation suivant, qu'ils ont appelé poursuite de base (Basis Pursuit, en anglais) :

$$(P1) \min_x \|x\|_1 \text{ tel que } y = Ax \quad (\text{III.7})$$

La question générale de savoir quand ' $l_0 = l_1$ ' est valable est la clé de l'acquisition comprimée. Des conditions nécessaires et suffisantes ont été fournies, qui dépendent non seulement de la parcimonie du vecteur original x , mais aussi de l'incohérence de la matrice d'acquisition A .

III.6 Exemple d'acquisition et de reconstruction

Cet exemple utilise la transformée en cosinus discrète (DCT) comme base. Le signal généré est la somme de deux sinusoïdes

$$f(t) = \sin(1394\pi t) + \sin(3266\pi t) \quad (\text{III.8})$$

Le signal est échantillonné à 40000 Hz et sa durée est 0.25 seconde, résultant en un vecteur f de longueur $N=5000$. Le graphique supérieur de la Figure III.5 montre une partie de ce signal, ainsi que certains des $M = 500$ échantillons prélevés aléatoirement.

Le graphique inférieur montre les coefficients c , obtenus en prenant la transformée en cosinus discrète inverse de f , avec deux pointes aux fréquences appropriées. Parce que les deux fréquences sont incommensurables, ce signal ne tombe pas exactement dans l'espace couvert par les fonctions de base DCT, et il y a donc quelques dizaines de coefficients significatifs non nuls; Le signal compressé est un vecteur b de M échantillons prélevés aléatoirement du signal original.

Le système linéaire résultant, $Ax = b$, est $M \times N = 500 \times 5000$. Il y a 10 fois plus d'inconnues que d'équations.

Pour reconstruire le signal, nous devons trouver la solution de l'équation $Ax = b$ qui minimise la norme l_1 , de x . Il s'agit d'un problème d'optimisation non linéaire, et plusieurs programmes basés sur MATLAB sont disponibles pour le résoudre. Le programme l_1 -magic, écrit par Justin Romberg et Emmanuel Candès [9] a été utilisé dans cet exemple. Le graphique supérieur de la Figure III.6 montre la solution résultante x .

On voit qu'il a relativement peu de grandes composantes et qu'il ressemble beaucoup au DCT inverse du signal d'origine. De plus, la transformée en cosinus discrète de x , illustrée dans le graphique inférieur, ressemble fortement au signal original.

À titre de comparaison, la Figure III.7 montre la solution traditionnelle obtenue en minimisant la norme l_2 [41].

b =échantillons prélevés aléatoirement.

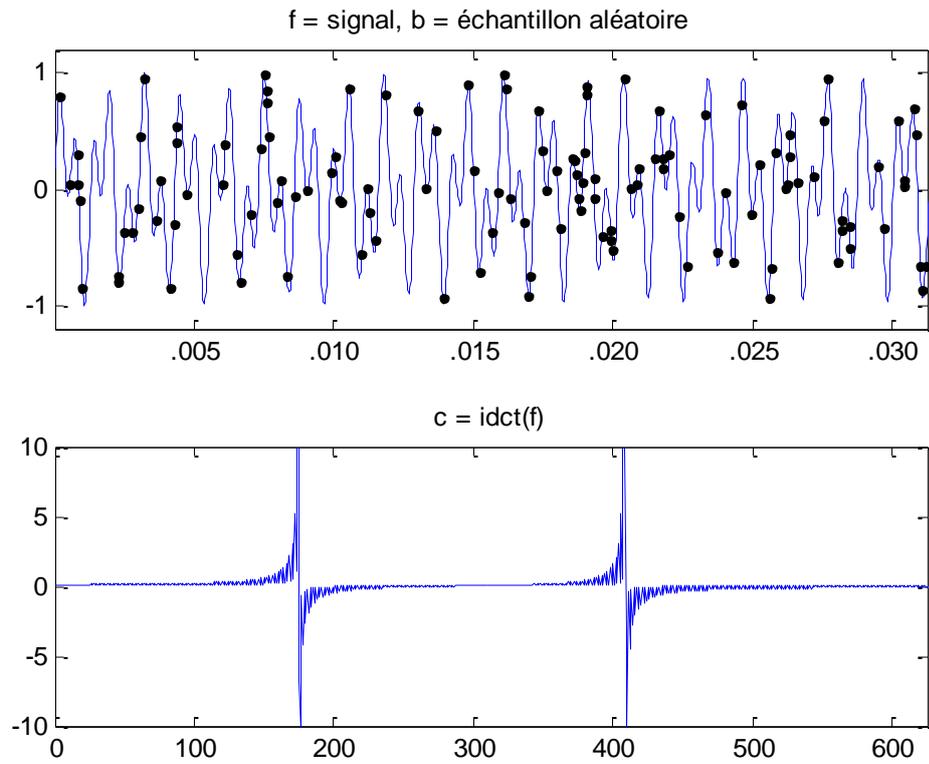


Figure III.5 : En haut : Échantillons prélevés aléatoirement du signal original. En bas : la transformée en cosinus discrète inverse du signal.

$x = \text{solution } l_1, Ax = b$

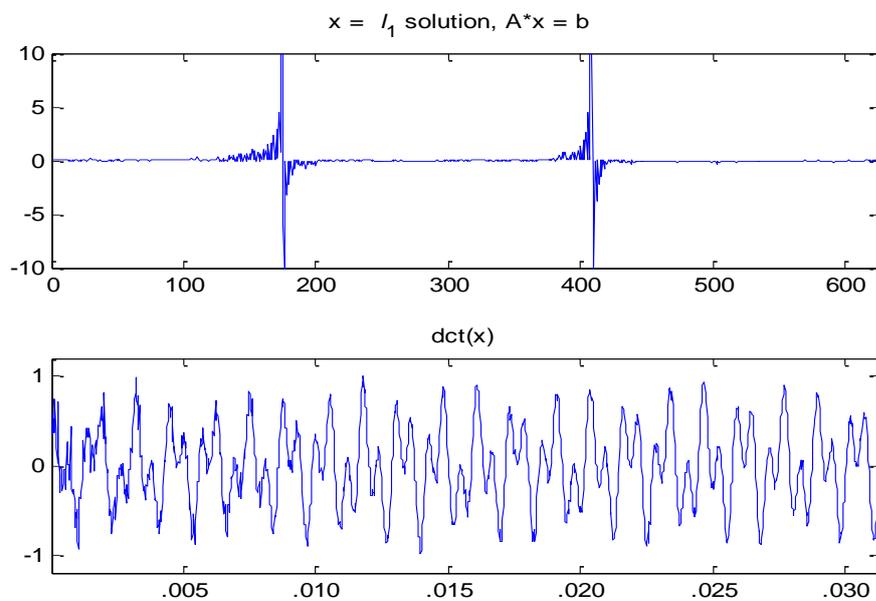


Figure III.6 : Solution de l'équation $Ax = b$ obtenue par minimisation l_1 .

$y = \text{solution } l_2, Ay = b.$

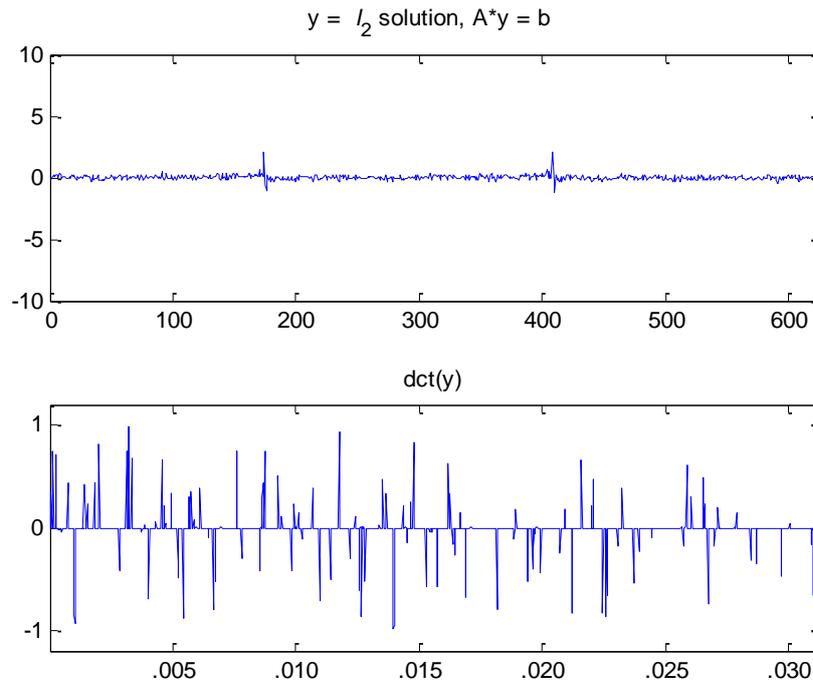


Figure III.7 : Solution de l'équation $Ay = b$ obtenue par minimisation l_2 .

III.7 Conclusion

Dans ce chapitre, nous avons présenté brièvement le principe de l'acquisition comprimée ainsi que les propriétés que la matrice de mesure devrait satisfaire. L'acquisition comprimée permet de reconstruire un signal parcimonieux dans un domaine donné en prélevant des échantillons à une fréquence inférieure à celle de Nyquist. De même, nous avons présenté un exemple d'illustration de l'acquisition comprimée et de la reconstruction en utilisant l'algorithme d'optimisation l_1 -magic.

chapitre IV :

*Acquisition comprimée du signal de parole
pour la télé-médecine*

IV.1 Introduction

Le signal de parole voisée est parcimonieux dans le domaine fréquentiel. Dans ce chapitre, la propriété de parcimonie est exploitée pour compresser le signal pour les applications de la télémédecine à de l'évaluation des troubles de la voix. Le signal reçu par le médecin est reconstruit pour calculer les indices acoustiques servant de descripteurs des troubles de la voix. Les indices acoustiques calculés reflètent les perturbations de la période fondamentale (jitter), de l'amplitude (shimmer) et la présence du bruit (HNR : Harmonics-to-noise ratio) et des dyspériodicités (SDR : signal-to-dysperiodicity ratio) dans le signal. Les performances de l'acquisition comprimée sont analysées en effectuant une analyse de corrélation des indices acoustiques avec les scores de l'évaluation auditive (grade G) et comparées à celles du signal original ainsi que le signal MP3.

Les indices acoustiques jitter, shimmer et HNR sont calculés avec le logiciel Praat [42]. L'indice SDR ainsi que la méthode de l'acquisition comprimée sont implémentés avec MATLAB. La compression MP3 est effectuée en utilisant un logiciel de conversion Freemake Audio Converter [43].

IV.2 Base des données

La base de données utilisée pour analyser les performances de l'acquisition comprimée comprend des voyelles [a] soutenues produites par 79 locuteurs espagnols normophoniques et pathologiques de différents âges [44]. L'évaluation auditive est effectuée par trois experts expérimentés en analyse de la qualité vocale. Les signaux vocaux ont été numérisés à 16 bits avec un taux d'échantillonnage de 44100 Hz. Les locuteurs devaient maintenir la voyelle espagnole /a/ pendant 4 secondes. Ensuite, les premiers et derniers segments de 0,5 seconde ont été coupés et la partie médiane de 3 secondes a été sélectionnée pour l'analyse acoustique. Pour étiqueter ces enregistrements, les évaluateurs ont utilisé l'échelle de perception Grade Roughness-Breathiness (GRB), qui est une version modifiée de l'échelle originale GRBAS. Dans ce travail, la performance de l'acquisition comprimée est analysée en faisant une analyse de corrélation des indices acoustiques avec le grade G représentant le degré d'enrouement global de la voix.

IV.3 Compression et reconstruction du signal de parole par acquisition comprimée

IV.3.1 Méthode

Le signal est divisé en trames de longueurs N . L'algorithme d'acquisition comprimée et de la méthode de compression basée sur acquisition comprimée consiste à extraire M

mesures linéaires ($M \ll N$) dans un domaine où le signal est parcimonieux. Dans ce travail, la transformée en cosinus discrète a été utilisée comme domaine de parcimonie. L'extraction des mesures est réalisée en utilisant une matrice de mesure Φ aléatoire de distribution gaussienne. Le signal compressé est alors transmis en utilisant les moyens de communications existantes (Mobile, Internet, ...). Le signal est reconstruit à partir du signal compressé en utilisant la minimisation l_1 et le programme d'optimisation l_1 -magic [41]. Les indices acoustiques sont alors calculés à partir du signal reconstruit.

Le principe de la méthode de compression et reconstruction du signal de parole par acquisition comprimée est illustré par la Figure IV.1. Les différentes étapes de l'algorithme peuvent être résumées comme suit :

1. Décomposer le signal original en trames de N échantillons.
2. Appliquer l'algorithme de l'acquisition comprimée aux différentes trames en utilisant une matrice de mesure aléatoire, la conception de la matrice de mesure doit permettre de retrouver les signaux les plus parcimonieux.
3. Transformer en cosinus discrète (DCT) Ψ .
4. Transmettre du signal compressé (sur mobile, sur internet, etc.).
5. Reconstruire du signal original et application de la minimisation l_1 .
6. Transformer cosinus discrète inverse.
7. Calculer des indices acoustiques.

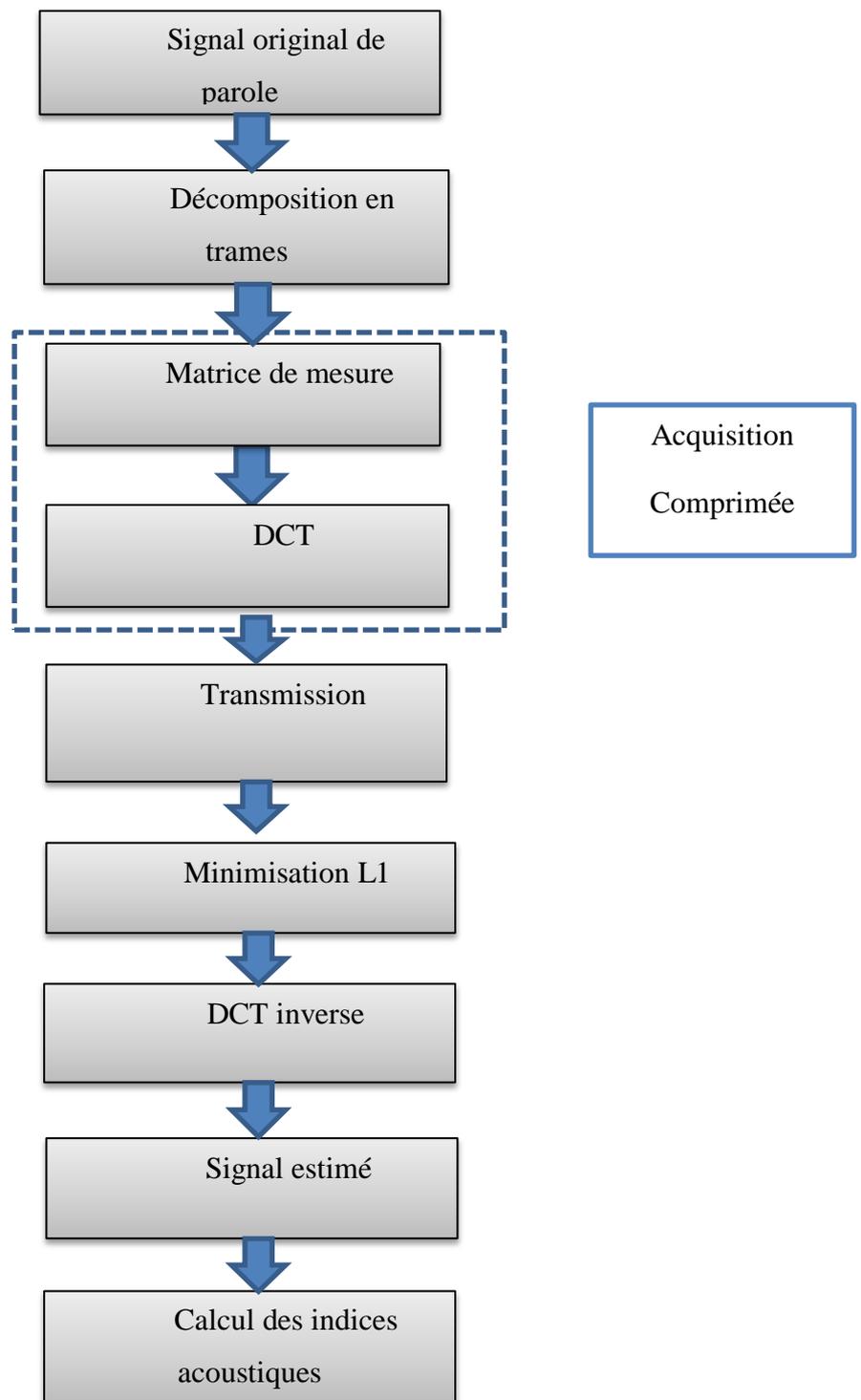
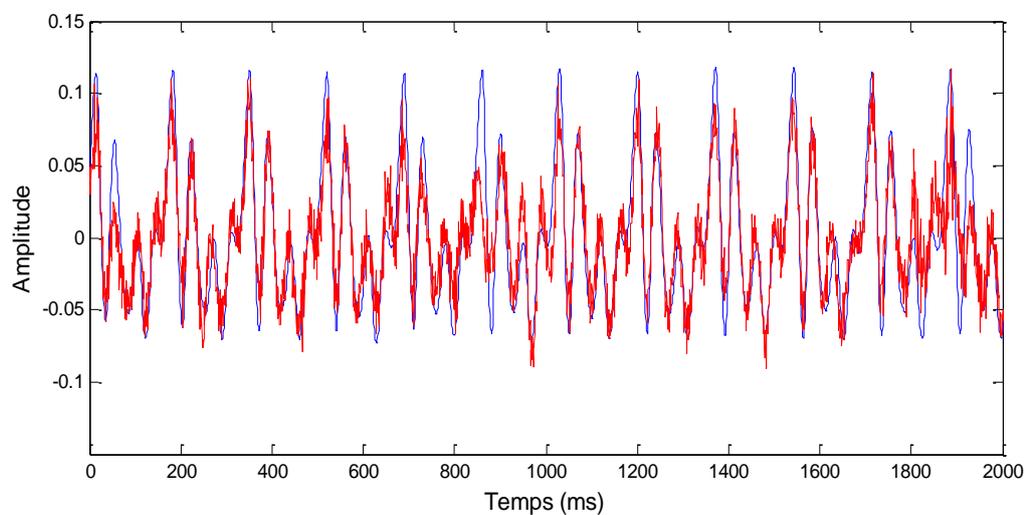


Figure IV.1: Schéma bloc de la méthode de compression/reconstruction du signal de parole par acquisition comprimée.

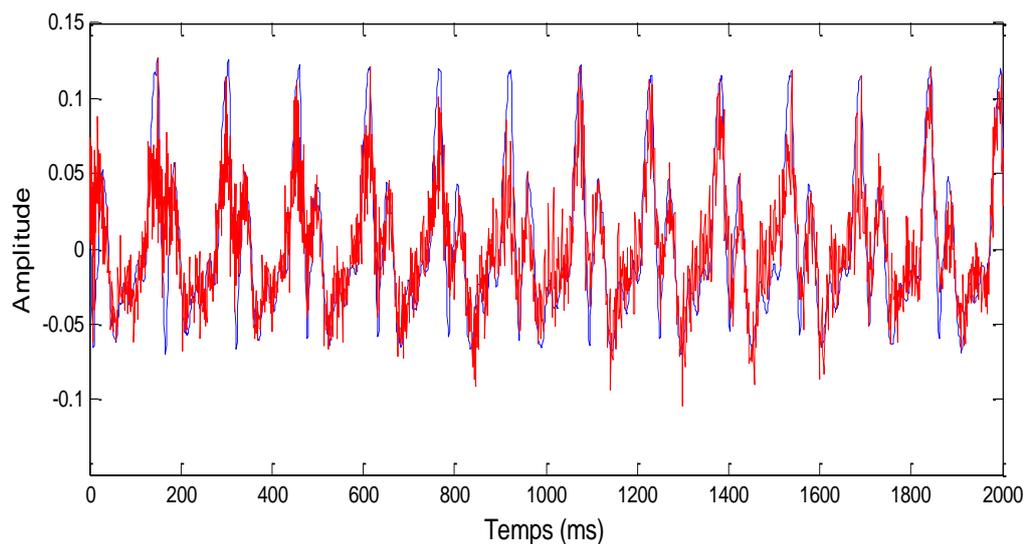
IV.3.2 Exemple d'illustration

Comme exemple d'illustration, les signaux acoustiques correspondants à la voyelle [a] produite par un locuteur normophonique (score 0) et un locuteur dysphonique (score 2) ont été compressés et reconstruits en utilisant la technique d'acquisition comprimée. La

Figure IV.2 montre les signaux reconstruits superposés au signal original. La longueur de trame est fixée 20 ms le nombre de mesurés prises aléatoirement $M=0.1N$, où N représente la longueur de la trame. Les valeurs des indices acoustiques jitter, shimmer et HNR calculés à partir du signal original et le signal reconstruit par acquisition comprimée sont données dans le tableau IV.1.



(a)



(b)

Figure IV.2: Signal reconstruit par acquisition comprimée (en rouge) superposé au signal original (en bleu). (a) score 0. (b) score 2.

Tableau IV.1 : Valeurs des indices acoustiques jitter, shimmer et HNR calculés à partir du signal reconstruit et du signal original pour un locuteur normophonique (score 0) et un locuteur dysphonique (score 2).

	Jitter		Shimmer		HNR	
	original	Estimé par AC	original	Estimé par AC	original	Estimé par AC
Score 0	0.253	1.629	1.166	14.224	27.184	6.365
Score 2	0.671	1.920	3.006	14.588	21.173	7.034

IV.4 Résultats et discussion

Le signal est divisé en trames de longueurs N et la technique d'acquisition comprimée est appliquée à chaque trame. Afin d'étudier l'effet de la longueur de trame et le taux de compression sur la performance de l'analyse acoustique du signal reconstruit en termes de corrélation des valeurs l'indice acoustique avec les scores de l'évaluation auditive (grade G), l'acquisition comprimée du signal a été effectuée pour différentes valeurs de la longueur de trame N et du nombre de mesures M prises aléatoirement. La longueur de trame N est de 10 ms, 20 ms et 40 ms pour les indices acoustiques jitter, shimmer et HNR et de 10 ms, 20 ms et 30 ms pour le SDR. Pour chaque valeur de N , les indices acoustiques sont calculés pour $M=0.1 N$, $M=0.2N$ et $M=0.3N$. Pour comparaison, on calcule aussi la corrélation des scores avec les indices acoustiques obtenus à partir du signal original et du signal MP3.

La Figure IV.3 montre l'évolution de la corrélation des valeurs du jitter avec les scores de l'évaluation auditive. On observe que le jitter calculé à partir du signal original est faiblement corrélé à l'évaluation auditive (corrélation de 0.44). La corrélation s'améliore légèrement lorsque l'indice acoustique est calculé à partir du signal MP3 (corrélation de 0.48) ou à partir du signal reconstruit par acquisition comprimée pour $M=0.3N$ (corrélation de 0.47) et ce quelle que soit la longueur de trame. L'acquisition comprimée permet d'améliorer la corrélation du jitter avec les scores de l'évaluation auditive si la longueur de trame est fixée à 40 ms et $M=0.1N$ (corrélation de 0.58).

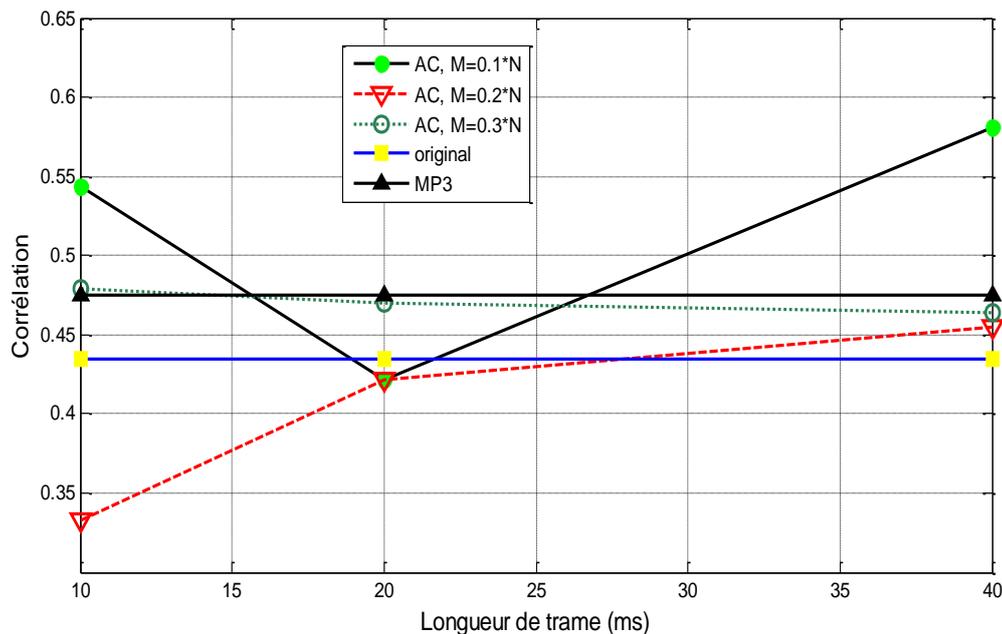


Figure IV.3 : Corrélation entre les scores de l'évaluation auditive et les valeurs du jitter calculé à partir du signal reconstruit par acquisition comprimée et à partir du signal original et du signal MP3.

La Figure IV.4 montre l'évolution de la corrélation des valeurs du shimmer avec les scores de l'évaluation auditive. Bien que la corrélation des scores de l'évaluation auditive avec les valeurs du shimmer calculé à partir du signal original soit modérée (corrélation de 0.62), elle est supérieure à celle obtenue à partir du signal reconstruit par acquisition comprimée et à partir du signal MP3. La plus grande corrélation obtenue à partir du signal reconstruit par acquisition comprimée est de 0.51 lorsque la longueur de trame est fixée à 40 ms et $M=0.3N$.

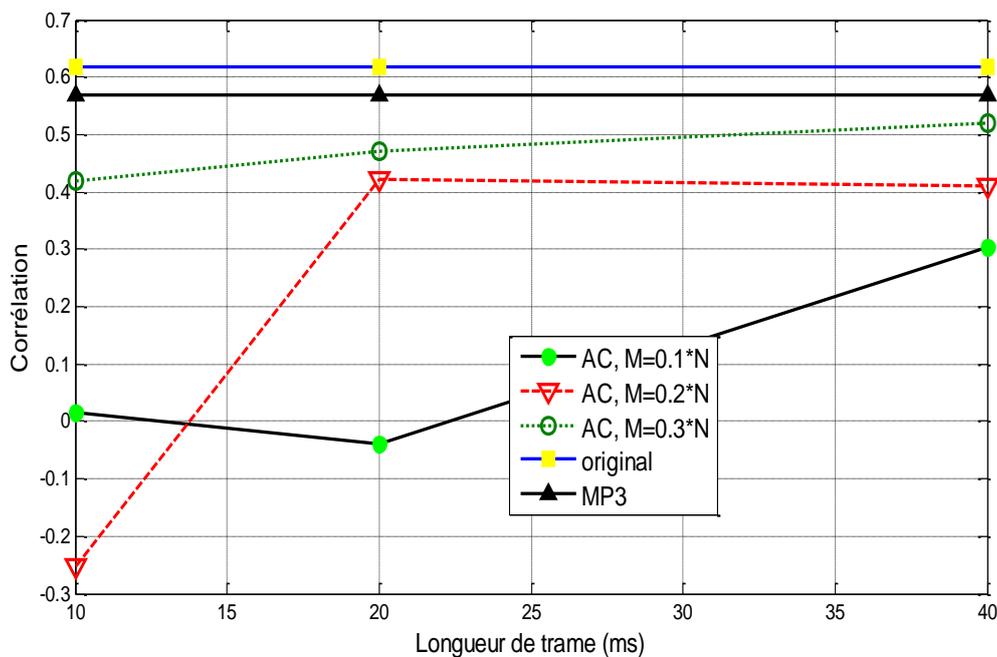


Figure IV.4 : Corrélation entre les scores de l'évaluation auditive et les valeurs du shimmer calculé à partir du signal reconstruit par acquisition comprimée et à partir du signal original et du signal MP3.

La Figure IV.5 montre l'évolution de la corrélation des valeurs du HNR avec les scores de l'évaluation auditive. On observe une faible corrélation entre les scores de l'évaluation auditive et les valeurs du HNR calculé à partir du signal original (corrélation de -0.65). La compression MP3 résulte en des valeurs identiques à celles obtenues à partir du signal original alors que l'acquisition comprimée conduit à une dégradation de la performance de l'indice acoustique en termes de corrélation avec les scores de l'évaluation auditive.

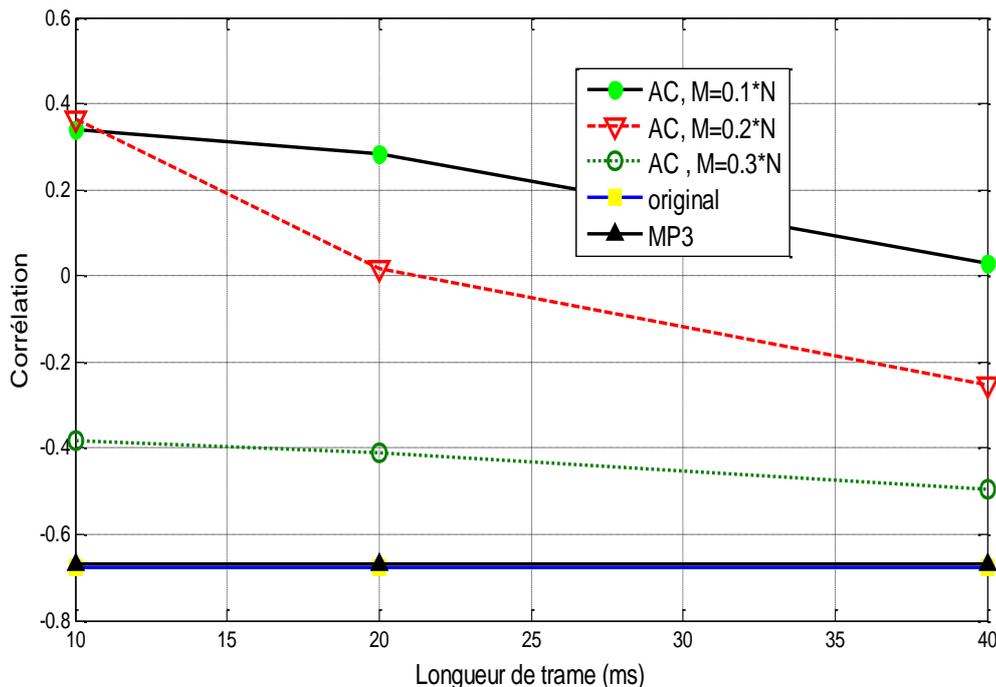


Figure IV.5: Corrélation entre les scores de l'évaluation auditive et les valeurs du HNR calculé à partir du signal reconstruit par acquisition comprimée et à partir du signal original et du signal MP3.

La Figure IV.6 montre l'évolution de la corrélation des valeurs du SDR avec les scores de l'évaluation auditive. La corrélation des scores de l'évaluation auditive avec les valeurs du SDR calculé à partir du signal original est de -0.66. L'application de l'acquisition comprimée améliore la performance de l'indice acoustique si la longueur de trame est fixée à 10 ms même pour un faible nombre d'échantillons. Cette corrélation est de -0.7 si $M=0.1N$. La compression MP3 résulte en une dégradation de la performance de l'indice acoustique en termes de corrélation avec les scores de l'évaluation perceptive.

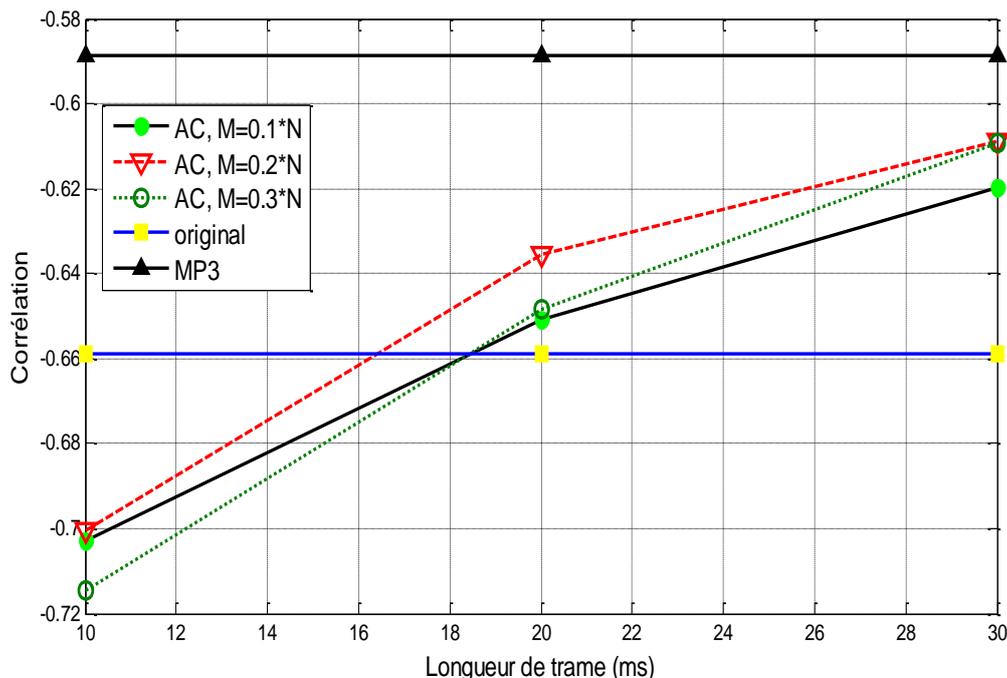


Figure IV.6: Corrélation entre les scores de l'évaluation auditive et les valeurs du SDR calculé à partir du signal reconstruit par acquisition comprimée et à partir du signal original et du signal MP3.

IV.5 Conclusion

Les résultats expérimentaux obtenus sur la base de données des voyelles [a] montrent que le calcul de l'indice acoustique SDR à partir du signal reconstruit par acquisition comprimée conduit à une plus forte corrélation des valeurs de l'indice avec les scores de l'évaluation auditive comparé à celui calculé à partir du signal original. Pour une longueur de trame à 10 ms, cette corrélation est de -0.7 en prenant $M=0.1N$. La compression MP3 dégrade la performance du SDR en termes de corrélation avec les scores de l'évaluation auditive. Pour le jitter, moyennant un choix adéquat de la longueur de trame, il est possible d'avoir une corrélation des scores de l'évaluation auditive supérieure à celle obtenue à partir du signal original et MP3. Toutefois, ces résultats restent à confirmer en utilisant d'autres bases de données.

Conclusion générale

Conclusion générale

La télémédecine offre un moyen pour les patients d'être examinés et traités, tandis que le fournisseur de soins, de santé et le patient se trouve à différents endroits physiques. Les applications de télémédecine à l'analyse et l'évaluation des troubles de la voix sont basées sur les technologies existantes pour la compression et la transmission du signal de parole signal développé pour les communications cellulaires et Internet.

L'acquisition comprimée a été proposée dans le but d'évaluer les troubles de la voix dans un système de télémédecine. L'acquisition comprimée est une méthode d'acquisition et de compression simultanées des signaux. Elle permet la reconstruction des signaux ayant une représentation parcimonieuse dans certaines bases en utilisant un nombre d'échantillons beaucoup plus faible que celui déterminé par le théorème de Shannon-Nyquist.

La méthode d'acquisition et de compression simultanée des signaux de parole a été testée sur une base de données comprenant des voyelles [a] produites par des locuteurs espagnols normophoniques et dysphoniques. Le choix des voyelles vient du fait que les sons voisés sont pseudo-périodiques et par conséquent les signaux sont parcimonieux dans le domaine fréquentiel.

L'effet de compression a été investigué en comparant les indices acoustiques jitter, shimmer, HNR et SDR obtenus à partir du signal reconstruit à ceux calculés en utilisant le signal original et le signal compressé MP3. Les résultats expérimentaux montrent que pour l'indice acoustique SDR, l'acquisition comprimée présente des performances en termes de corrélation avec les scores de l'évaluation auditive meilleures que la compression MP3 et le signal original lui-même. Pour le jitter, moyennant un choix adéquat de la longueur de trame, il est possible d'avoir une corrélation des scores de l'évaluation auditive supérieure à celle obtenue à partir du signal original et MP3.

Comme future travail, l'acquisition comprimée peut être testée sur d'autres bases de données de plus grandes tailles, sur la parole continue. De même, comme extension à ce travail, la méthode peut être proposée pour la compression de la parole dysarthrique pour l'évaluation des troubles de la parole en télémédecine.

Références bibliographiques

Références bibliographiques

- [1] T. Martin, “La téléconsultation du médecin généraliste en zone rurale fragile : vers une nouvelle forme d’inégalité d’accès aux soins ?,” Mémoire de Master, Ethique en Santé, Université de Caen Normandie, France, 2021.
- [2] S. Sandoval et al, “The Effects of Speech Compression Algorithms on the Intelligibility of Two Individuals With Dysarthric Speech,” *American Journal of Speech-Language Pathology*, Vol. 28, 195–203, February 2019.
- [3] B. Homayoon, *Fundamentals of Speaker Recognition*, Springer, 2011.
- [4] F. Rudzicz, *Articulatory Knowledge in the Recognition of Dysarthric Speech*, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, No. 4, 2011.
- [5] C. Mertens, “Analysis of vocal tremor in normophonic and dysphonic speakers,” PhD thesis, Université Libre de Bruxelles, 2015.
- [6] G. Léothaud, *Théorie de la phonation*, cours de DEUG 2^{ème} année, P 20, 2004-2005.
- [7] D. Gargouri, “Contribution à l'estimation et à la poursuite des trajectoires de formants de parole,” Thèse, Ecole Nationale d'Ingénieurs de Sfax, Sfax, Tunisie, 2010.
- [8] <https://p.21-bal.com/law/6245/index.html>
- [9] J. H. Deller, J. G. Proakis and J. H. L. Hansen, *Discrete-time processing of speech signals*, Prentice-Hall, 1993.
- [10] Z. Gamache et I. Boudjir, “Analyse en composantes principales du spectrogramme du signal de parole : application à la détection de la dysathrie,” Mémoire de Master, Systèmes des Télécommunications, Département d’Electronique, Université Mohammed Seddik BENYAHIA-Jijel, Algérie, 2017.
- [11] R. Boite et M. Kunt, *Traitement de la parole*, Presses Polytechniques Romandes, Lausanne, 1987.
- [12] Calliope, *la Parole et son Traitement Automatique*, Masson, Paris, 1989.
- [13] J.L Roulin, *Psychologie cognitive*, 2^{ème} édition, Bréal, Rome, 2016.
- [14] C. Mertz, “Étude des corrélations entre les évaluations subjectives et objectives de la voix chez des patients présentant une paralysie laryngée bilatérale,” Mémoire de Certificat de Capacité en Orthophonie, Département d’Orthophonie, UFR Santé, Université de Rouen Normandie, France 2018.

Références bibliographiques

- [15] E. Sicard, A. Menin-Sicard, “Etude Bibliographique sur les Indicateurs de Pathologie de la Voix,” Rapport de recherche, INSA Toulouse, France, 2013.
- [16] F. Leclère, “Estimation du rapport signal sur bruit en sortie d’aide auditive; mise œuvre, Application,” Mémoire de Master, Etat d’Audioprothèse, 2014
- [17] A. Kacha, F. Grenez, and J. Schoentgen, “Estimation of dysperiodicities in disordered speech,” *Speech Comm*, Vol. 48, No 10, pp. 1365-1378, 2006.
- [18] S. Stankovic´, I. Orovic´, and E. Sejdic´, *Multimedia Signals and Systems Basic and Advanced Algorithms for Signal Processing*, 2nd edition, 2016.
- [19] B. Magnus, “Wireless communication in telemedicine using Bluetooth and IEEE 802.11b,” Master’s Thesis, Department of Information Technology Uppsala University, 2001.
- [20] L. Bajolle, “E-médecine : Amélioration, Optimisation et Humanisation de la médecine de ville par l’usage de l’internet et des nouvelles technologies,” Thèse de doctorat en médecine de l’Université Joseph Fourier, Grenoble, janvier 2002.
- [21] C. Suarez, “La télémédecine : quelle légitimité d’une innovation radicale pour les professionnels de santé ?,” *Revue de l’Institut de Recherches Économiques et Sociales (IRES)*, vol. 39, 2002.
- [22] S. Pierre, A. Dominique, “La place de la télémédecine dans l’organisation des soins,” rapport de la Direction Générale de l’Hospitalisation et de l’Organisation des Soins pour le Ministère de la santé et des Sports, France, 2008.
- [23] F. Kohler, *Télémédecine, télésanté, eSanté*, université Nancy, sante, France, 2010.
- [24] M. Cauville, “ Diagnostic, soins et prévention par la télémédecine: explications de J. Demongeot,” *Sciences et Technologies*, Vol. 2, pp. 32–34, 1999.
- [25] A. Nemo, “La télémédecine : Faire voyager les informations plutôt que le malade,” *Journal du Téléphone*, pp. 4, 1994.
- [26] S. Krim, Y. Benmansour, “Teleimagerie medicale mobile,” Mémoire de Master, Université de Abou bekr Belkaid, 2012.
- [27] D. Florence, “Fusion de données multicapteurs pour un système de télésurveillance médicale de personnes à domicile,” Thèse de Doctorat, Université Joseph-Fourier-Grenoble I, 2004.

Références bibliographiques

- [28] S.Rerbal, “ traitement numérique du signal physiologique en télémédecine,” Thèse de Doctorat, Université Aboubekrbelkaid, Tlemcen, Algérie, 2014.
- [29] D.G. Theodoros, A.J. Hill, Russell, T. Ward, E.C. and Wootton, R. (2008). Assessing acquired language disorders in adults via the Internet. *Telemedicine and e-Health*, 14, 552-559.
- [30] S. Bouasli, A. Noumeri, “compression et codage de la parole par la transformée KLT automatisée,” Mémoire de Master, Systèmes de Télécommunications, Université Djilali Bounaama Khemis Miliana, 2016.
- [31] J. Raverdino, J M. Guatteri, *Techniques multimédia pour le son*.
- [32] J. Hernandez, “algorithmes d’acquisition, compression et restitution de la parole à vitesse variable. Etude et mise en place,” Ecole Nationale Supérieure de l’Electronique et de ses Applications Cergy-Pontoise, Paris, 1995.
- [33] [https://olczyk.pagesperso-orange.fr/physique-chimie/Ens-Sc-1ere/EnSc14.2%20Numérisation fichiers/MP3.pdf](https://olczyk.pagesperso-orange.fr/physique-chimie/Ens-Sc-1ere/EnSc14.2%20Numérisation%20fichiers/MP3.pdf)
- [34] H.Haneche, “Compressed Sensing in Mobile Systems,” Doctoral Thesis, Telecommunications and Information Processing, Electronics Department, University of Science and Technology Houari Boumediene, Algérie, 2020.
- [35] D.S. Taubman and M.W. Marcellin, *JPEG 2000 Image Compression Fundamentals, Standards and Practice*. Norwell, MA: Kluwer, 2001.
- [36] J. E. Candès and M. B. Wakin, “An Introduction To Compressive Sampling,” *IEEE SIGNAL PROCESSING MAGAZINE*, 2008.
- [37] Brunton, S. L., and Kutz, J. N. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2019.
- [38] S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM J. Sci. Comput.*, 20:33–61, 1998.
- [39] S. Muthukrishnan, *Data Streams: Algorithms and Applications*. Now Publishers, Boston, MA, 2005.
- [40] G.Kutyniok, *Theory and Applications of Compressed Sensing*, October 15, 2018.
- [41] M. Cleve, “Magic” Reconstruction: Compressed Sensing, MathWorks, 2010.
- [42] P. Boersma, D. Weenink, "Praat: doing phonetics by computer" [Computer program]. Version 6.0.46, <http://www.praat.org/>, 3 February 2019.
- [43] <https://download.freemake.net/Reserved/FreemakeAudioConverterSetup.exe>

Références bibliographiques

- [44] R. M. Bermúdez de Alvear et al., “A database and digital signal processing framework for the perceptual analysis of voice quality,” 11th PAN European voice conference, August 31st-September 4th 2015, Firenze, Italy.

Résumé

La télé-médecine est définie comme une forme de pratique médicale à distance utilisant les technologies de l'information et de la communication. Les applications de télé-médecine pour l'analyse et l'évaluation des troubles de la voix sont basées sur les technologies existantes pour la compression et la transmission du signal de parole développé pour les communications cellulaires et Internet. Il est important donc de connaître l'effet des algorithmes de compression sur la parole pathologique.

Dans ce mémoire, on se propose d'utiliser l'acquisition comprimée comme alternative aux algorithmes conventionnels pour la compression des signaux de parole dans les applications de télé-médecine. L'acquisition comprimée est une méthode d'acquisition et de compression simultanées des signaux. Elle permet la reconstruction des signaux ayant une représentation parcimonieuse dans certaines bases en utilisant un nombre d'échantillons beaucoup plus faible que celui déterminé par le théorème de Shannon-Nyquist.

L'effet de compression est investigué en comparant les indices acoustiques obtenus à partir du signal reconstruit à ceux calculés en utilisant le signal original et le signal compressé MP3.

Mots clés : acquisition comprimée, télé-médecine, indices acoustique, compression du signal de parole, évaluation des troubles de la voix.

Abstract

Telemedicine is defined as a form of remote medical practice using information and communication technologies. The telemedicine applications for the analysis and evaluation of voice disorders are based on existing technologies of compression and transmission of the speech signal which is developed for cellular and Internet communications. So, it is important to know the effect of compression algorithms on pathological speech.

In this thesis, we propose to use compressed sensing as an alternative to conventional algorithms of compression of speech signals in telemedicine applications. Compressed sensing is a method of simultaneously acquiring and compressing signals. It allows the reconstruction of signals having a sparse representation in certain bases by using a number of samples much fewer than what is determined by the Shannon-Nyquist theorem.

The compression effect is investigated by comparing the acoustic indices obtained from the reconstructed signal to those calculated using the original signal and the compressed MP3 signal.

Key words: compressed sensing, telemedicine, acoustic indices, speech signal compression, voice disorders assessment.

ملخص

يُعرّف الطب عن بُعد بأنه شكل من أشكال الممارسة الطبية عن بُعد باستخدام تقنيات المعلومات والاتصالات. تعتمد تطبيقات التطبيب عن بُعد لتحليل وتقييم اضطرابات الصوت على التقنيات الحالية لضغط ونقل إشارة الكلام المطورة للاتصالات الخلوية والإنترنت ، لذلك من المهم معرفة تأثير خوارزميات الضغط على الكلام المرضي.

في هذه الرسالة، نقترح استخدام الاستحواذ المضغوط كبديل للخوارزميات التقليدية لضغط إشارات الكلام في تطبيقات التطبيب عن بُعد. الاستحواذ المضغوط هو طريقة للحصول على الإشارات وضغطها في وقت واحد. يسمح الاستحواذ المضغوط بإعادة بناء الإشارات ذات التمثيل المنتثر في قواعد معينة باستخدام عدد من العينات أقل بكثير من تلك التي تحددها نظرية Shannon-Nyquist .

يتم التحقق من تأثير الانضغاط من خلال مقارنة المؤشرات الصوتية التي تم الحصول عليها من الإشارة المعاد بناؤها مع تلك المحسوبة باستخدام الإشارة الأصلية وإشارة MP3 المضغوطة.

الكلمات المفتاحية: الاستحواذ المضغوط ، التطبيب عن بُعد ، مؤشر صوتي ، ضغط إشارة الكلام ، تقييم اضطراب الصوت.