

Table des matières

Liste des figures :	iii
Liste des tableaux :	iv
Liste des Abréviations :	iv
REMERCIEMENTS	v
INTRODUCTION	1
Chapitre I. Le principe de la reconnaissance vocale	4
I.A. GENERALITES	5
I.A.1. L'appareil phonatoire	6
I.A.2. Signal de la parole	7
I.A.3. Prosodie.....	11
I.A.4. L'information vocale [6] [14]	12
I.B. METHODE DE RECONNAISSANCE VOCALE	12
I.B.1. Le prétraitement du signal	14
I.B.2. La comparaison Dynamique ou DTW	34
Chapitre II. Réalisation du didacticiel nommé « PC Helper ».....	36
II.A. PROGRAMMATION.....	37
II.B. LES INTERFACES DE « PC Helper ».....	38
II.B.1. Ecran de démarrage	38
II.B.2. Ecran de connexion	39
II.B.3. Ecran d'accueil.....	40
II.B.4. Ecran détail.....	41
II.B.5. Ecran Vidéo.....	42
II.B.6. Menu	43
II.B.7. Ecran EXERCICE.....	44
II.B.8. Ecran COMMANDE	45
II.B.9. Ecran Favoris	46
II.B.10. Ecran INFORMATION	47
II.B.11. Ecran Checking Internet	48
II.B.12. Ecran MICRO.....	49
Chapitre III. Les points forts de l'intégration de la reconnaissance vocale dans le secteur de l'éducation.....	50

III.A. De réels avantages pour tous les étudiants	51
III.A.1. Rapidité	51
III.A.2. Prendre de notes :	54
III.B. Accessibilité des étudiants souffrant d'un handicap physique	56
III.B.1. Utilisation de la tablette en mains libres	56
III.B.2. Une aide pédagogique pour les étudiants sourds et malentendants.	56
III.C. Technologie d'assistance à l'usage des étudiants souffrant de difficultés d'apprentissage.	57
III.C.1. Transformer les idées en textes écrits	58
III.C.2. Permettre une plus grande autonomie.....	58
CONCLUSION.....	60
Bibliographies :	62
Webographie :	63

Liste des figures :

Figure 1 Types de reconnaissance vocale.....	5
Figure 2 Les organes de la phonation.....	6
Figure 3 Estimation des largeurs de bande des formants [6].....	7
Figure 4 Signal acoustique s [11].....	9
Figure 5 Système de reconnaissance vocale.....	13
Figure 6 Schéma bloc du prétraitement [12].....	14
Figure 7 Classification morphologique des signaux.....	15
Figure 8 Echantillonnage idéal.....	16
Figure 9 Composition spectral d'un signal.....	17
Figure 10 Aspect fréquentiel du sous – échantillonnage	17
Figure 11 Echantillonneur-bloqueur d'un signal analogique	18
Figure 12 Description de la fonction blocage.....	19
Figure 13 Chaîne de numération.....	20
Figure 14 Modélisation de quantification	20
Figure 15 Quantification Uniforme et bruit de quantification.....	21
Figure 16 Répartition des fréquences discrètes	27
Figure 17 Représentation temporelle du mot "physique" (en haut) ainsi qu'un extrait de 30 ms pendant la prononciation de [Ph] (en bas).....	27
Figure 18 Courbe de la fonction de pondération de Hamming sur 128 points [16]	29
Figure 19 Représentation d'un signal sinusoïdal non-pondéré puis pondéré par la fonction de Hamming	29
Figure 20 Banc de filtres espacés selon l'échelle de Mel (ici $F_e=8000$ Hz) [11]	30
Figure 21 Représentation de la matrice $N \times K$ des coefficients du banc de filtre (N: nombre d'échantillons par frame, K nombre de filtres souhaités) [11].....	33
Figure 22 Chemin idéal entre 2 spectres [6].....	35
Figure 23 Ecran de démarrage.....	38
Figure 24 Ecran de connexion.....	39
Figure 25 Ecran d'accueil	40
Figure 26 Ecran détail.....	41
Figure 27 Ecran vidéo.....	42
Figure 28 Ecran menu.....	43
Figure 29 Ecran exercice	44
Figure 30 Ecran commande	45
Figure 31 Ecran Favoris.....	46
Figure 32 Ecran information.....	47
Figure 33 Ecran checking internet	48
Figure 34 Ecran Micro	49
Figure 35 Ecran d'exercice d'application.....	52
Figure 36 Ecrans d'application favoris 1	53
Figure 37 écrans d'application favoris 2.....	53
Figure 38 Ecran de note 1	54
Figure 39 Ecran de note 2	55

Liste des tableaux :

Tableau 1 : Estimations des largeurs de bande des formants, obtenues par différents expérimentateurs [6]	7
Tableau 2 Les phonèmes de la langue française.....	11
Tableau 3 Echelle de fréquence Mel.....	31
Tableau 4 Spécifications techniques du logiciel « PC Helper ».....	37

Liste des Abréviations :

dB : décibel

TF : Transformation de Fourier

TFN : Transformation de Fourier Numérique

TFN : Transformation de Fourier Rapide

FFT : Fast Fourier Transform

TFD : Transformation de Fourier Discrète

DCB : Décimal Codé Binaire

DTW: Dynamic Time Warping

CAN : Convertisseur Analogique-Numérique

JDK : Java Development Kit

X(t) : signal analogique

REMERCIEMENTS

Nous sommes profondément reconnaissants envers Dieu Tout Puissant de nous avoir donné la force, la santé, l'intelligence et le courage pour mener à terme nos cinq années d'études à l'Ecole Normale Supérieure ainsi que pour la réalisation de ce mémoire de fin d'études. A lui seul l'honneur et la gloire. Merci beaucoup Seigneur.

Nous exprimons notre très haute considération et nos premiers mots de remerciements les plus sincères à Mr RASOLONDRAMANTRA HENRY Ph.D et Maître de conférences qui nous a fait l'honneur de présider la soutenance de ce mémoire.

Nous exprimons également notre entière reconnaissance à nos juges qui, en dépit de leurs multiples préoccupations, n'ont ménagé ni leurs temps, ni leurs efforts pour avoir veillé au bon achèvement de ce travail en lisant, en corrigeant et en proposant des remarques pour améliorer notre recherche. Il s'agit de Mr RASOANAIVO René, Ph.D, Maître de conférences et ainsi que Mr RANDRIATEFISON Nirilalaina, Maître de conférences .Nous exprimons également notre profonde gratitude à notre rapporteur Professeur ANDRIANARIMANANA Jean Claude Omer pour ses conseils, sa disponibilité, son dévouement et surtout pour ses critiques et suggestions qui nous ont été d'un apport considérable. La réalisation de ce mémoire n'aurait été possible sans votre bienveillant soutien Monsieur, ainsi merci infiniment.

Enfin, nous tenons à remercier nos parents, mon frère et ma sœur ainsi que toute la famille pour leur soutien effectif, financier et spirituel tout au long de la préparation de ce travail et durant nos 5 années d'étude à l'ENS. De plus, nous adressons nos vifs remerciements à tous les enseignants et personnels administratif auprès de l'Ecole Normale Supérieure, à tous nos collègues de la promotion FANATENANA, et les amis de près ou de loin.

« Merci à toutes et à tous »

INTRODUCTION

Apprendre les sciences physiques demande de la part des élèves la compréhension des phénomènes étudiés. De même, enseigner les sciences physiques vise à développer chez l'apprenant une capacité de savoir-faire et un esprit scientifique, c'est-à-dire, un esprit d'analyse, de critique et d'objectivité. Une question fondamentale se pose alors : comment faire pour atteindre ces objectifs sachant que beaucoup des élèves ont des difficultés sur le processus d'apprentissage ?

De nos jours, la majorité des écoles primaires et secondaires intègrent dans leur programme l'utilisation de la nouvelle technologie de communication et de l'information, elle apporte une nouvelle approche dans le processus de l'enseignement. Elle est aussi d'une grande aide pour les étudiants souffrant de handicaps physiques ou de difficultés d'apprentissage.

Les études et les recherches menées au sein du Centre d'Etudes et Recherches en Physique et Chimie (CER-PC) de l'Ecole Normale Supérieure d'Antananarivo sont axées sur la production des produits de technologie de l'Information et de Communication pour l'Enseignement (TICE) pour l'amélioration de la qualité de l'enseignement et de l'apprentissage des sciences Physiques.

C'est ainsi que, nous avons choisi dans le cadre de ce mémoire de fin d'études, le sujet de mémoire

« Technique d'enseignement via une application reconnaissance vocale »

Il s'agit d'une ressource numérique qui vise à

- Numériser le son (en permettant chez l'apprenant(e) de passer des idées au texte plus facilement).
- Aider les élèves souffrant de difficulté d'apprentissage (Limitant leur apprentissage en raison d'handicaps physiques, en raison de trouble d'apprentissage comme l'erreur de calcul, faute d'orthographe.....).
- Faciliter et améliorer l'enseignement/apprentissage de la physique chimie en utilisant une tablette comme outil didactique, afin d'amener les apprenant(e)s à construire eux-mêmes leurs connaissances.

Notre travail personnel est présenté en trois parties :

- La première partie concerne sur le principe de la reconnaissance vocale
- La deuxième partie expose la réalisation du didacticiel nommé « PC Helper »
- Et la dernière partie les points forts d'intégration de la reconnaissance vocale dans l'enseignement/apprentissage.

Chapitre I. Le principe de la reconnaissance vocale

I.A. GENERALITES

Il existe deux grands types de reconnaissance vocale (**figure1**) :

- ❖ La reconnaissance du locuteur ou « Speaker recognition » :

Le but de la reconnaissance du locuteur est de reconnaître qui parle et donc de reconnaître le locuteur et non pas le contenu.

- ❖ La reconnaissance de la parole ou « Speech recognition » :

La reconnaissance de la parole est utilisée pour reconnaître ce qui est dit. Dès lors, elle permet de transformer le discours/la voix en texte. Beaucoup de personnes pensent que la reconnaissance vocale et la reconnaissance de la parole sont similaires alors que ce n'est pas du tout le cas. Seul le contenu de la parole est reconnu par la reconnaissance de la parole.



Figure 1 Types de reconnaissance vocale

I.A.1. L'appareil phonatoire

Le résonateur de l'appareil phonatoire est composé de quatre cavités principales en série (**Figure2**) : le pharynx ou arrière gorge (1), les deux cavités buccales (2 et 3) délimitées par la langue et que l'on simplifiera à une seule et l'ajutage labiale (4) situé entre les dents et les lèvres. La cavité nasale, en "parallèle" sur l'ensemble série précédent, vient compléter ce résonateur.

La source de ce résonateur est en fait décomposable en deux émissions distinctes et d'origines différentes. Les cordes vocales, en fournissant un spectre riche en harmoniques, produisent les sons voisés. Le bruit d'écoulement de l'air en provenance des poumons, dont le spectre est similaire à un bruit blanc, crée les sons non voisés.

Les sons et la parole naissent de l'excitation d'un résonateur et sont formés par les ouvertures et les volumes de ce dernier qui varient très rapidement.

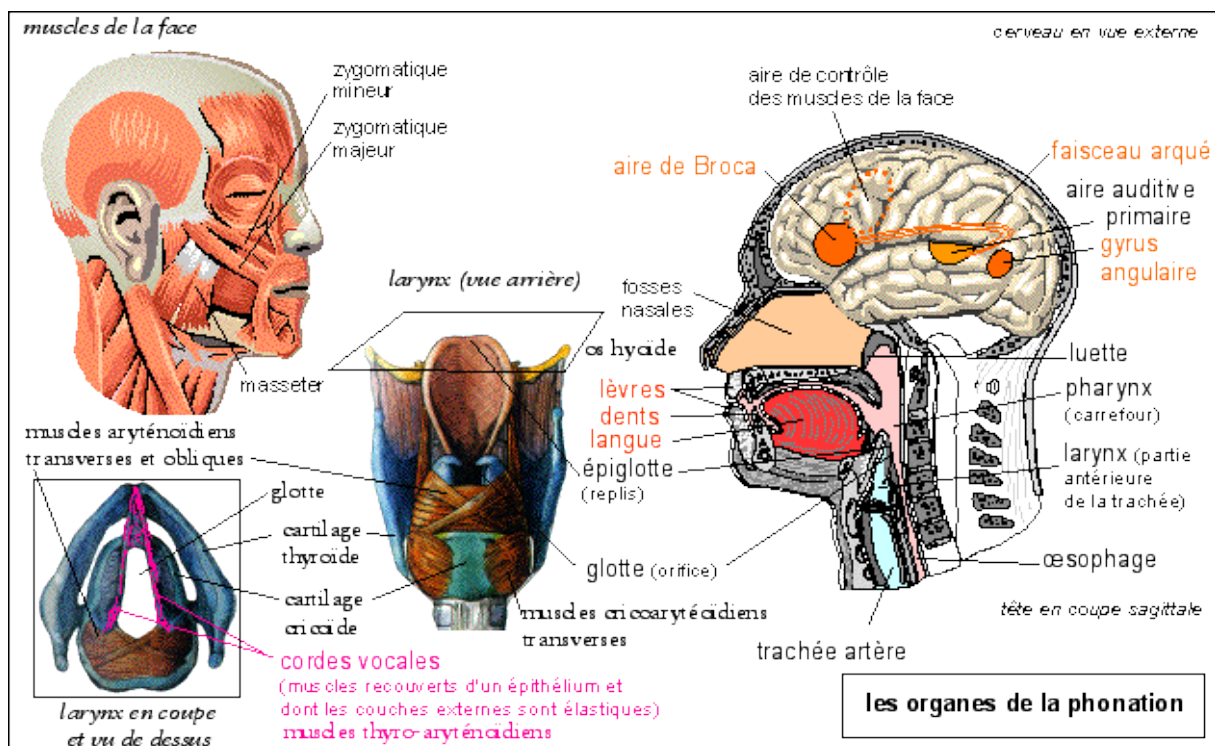


Figure 2 Les organes de la phonation

L'observation spectrale du conduit vocal laisse apparaître des pics de résonance, appelés « formants » (**figure 3**). Les affaiblissements constatés dans le spectre, nommés anti formants, sont introduits par les sons nasalisés.

Formants	Domaine de fréquence (H_z)	Domaine des largeurs de bande (H_z)
Formant 1	100 – 1 100	45 – 130
Formant 2	1 500 – 2 500	50 - 190
Formant 3	1 500 - 3 500	70 - 260

Tableau 1 : Estimations des largeurs de bande des formants, obtenues par différents expérimentateurs [6]

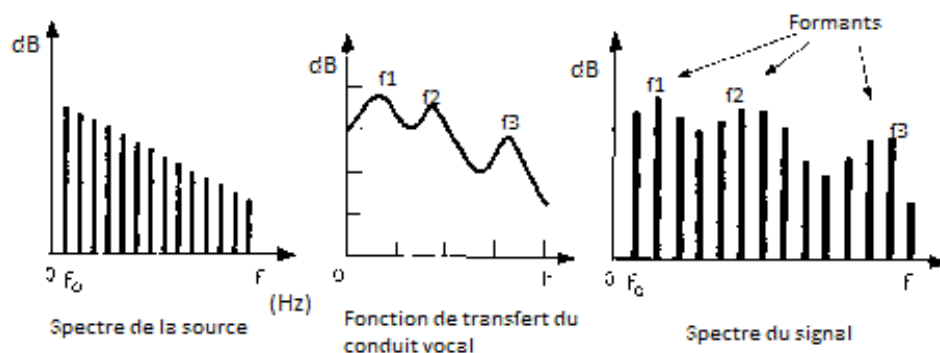


Figure 3 Estimation des largeurs de bande des formants [6]

I.A.2.Signal de la parole

Le signal vocal est caractérisé par :

- sa fréquence
- son intensité (ou le niveau sonore)
- son timbre (ou « la richesse » du signal)

I.A.2.a. Fréquence

La fréquence qui est l'inverse de la période **T**, est le nombre d'oscillations dans une seconde.

On peut aussi avoir la fréquence **F** à partir de la formule suivante :

$$\mathbf{F = V / \lambda}$$

Avec **V** la célérité ou la vitesse du son en **m/s** et **λ** la longueur d'onde en **m**.

➤ Fréquence fondamentale

Le signal de la parole comprend un son fondamental et des harmoniques dont les rapports de fréquences avec la fondamentale sont des quotients de nombres entiers. Toute vibration sonore peut être décomposée en une somme de fonctions sinusoïdales élémentaires dont les périodes plus courtes sont proportionnelles avec sa propre période. (C'est la décomposition en « série de Fourier »).

➤ Fréquence harmonique

C'est la fréquence multiple de la fréquence fondamentale **f₀**, c'est-à-dire **fn = nf₀** avec **n ∈ {2, 3, ...}**. Une harmonique correspond à une fonction trigonométrique sinusoïdale dont la fréquence est un multiple de la fréquence de la fonction périodique décomposée. La somme de toutes les harmoniques d'une fonction périodique reconstitue la fonction.

Prenons par exemple un signal acoustique **s(t)** qui est la superposition de trois sinusoïdales pures dont la fréquence fondamentale est **f₁ = 440Hz** et de fréquences harmoniques **f₂ = 880Hz, f₃ = 1320Hz**, d'équation

$$\mathbf{s(t) = \sin 2\pi f_1 t + \sin 2\pi f_2 t + \sin 2\pi f_3 t}$$

Les graphes de cette équation en fonction du temps et de la fréquence sont illustrés par la **(figure 4)**

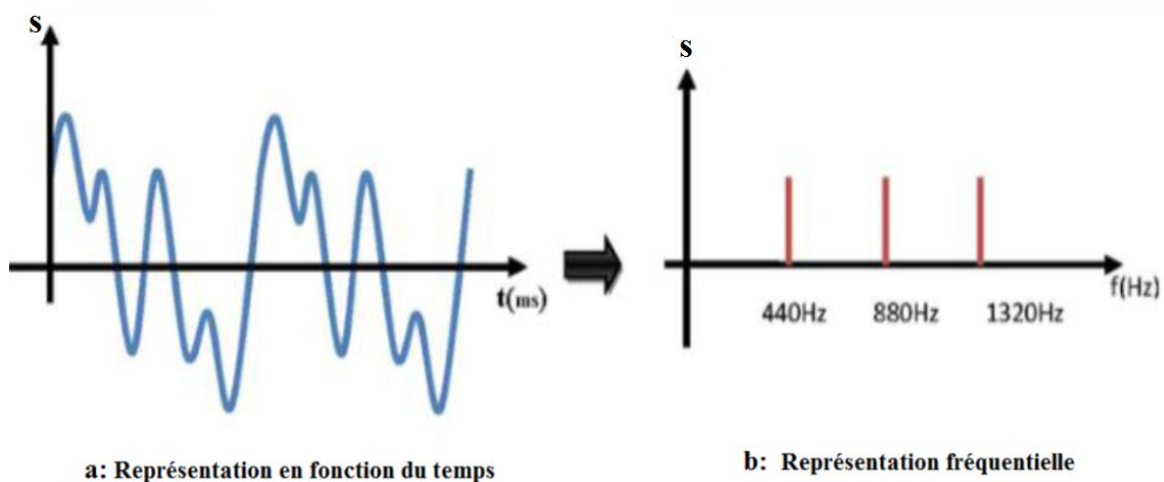


Figure 4 Signal acoustique s [11]

I.A.2.b. Intensité

Le son est une onde qui se propage de façon omnidirectionnelle. L'énergie acoustique E (en joule) produite par la source sonore est répartie sur une surface sphérique de plus en plus grande au fur et à mesure de la propagation de l'onde sonore. L'intensité notée I , est une fonction dépendante de la puissance P de l'émetteur et la distance notée r , qui sépare la source et le lieu d'écoute. [13]

L'intensité est donnée par la formule :

$$I = \frac{P}{4\pi r^2}$$

, avec P est en W et r en $m \Rightarrow I$ en W/m^2 .

Elle caractérise aussi le volume du signal mais dépend de l'amplitude du signal considéré. En acoustique, elle s'exprime en décibel en raison de deux facteurs :

- Les valeurs obtenues sont faciles à manipuler (elles ne sont trop grandes ni trop petites)
- La perception humaine sur l'intensité sonore se fait de façon logarithmique.
L'intensité acoustique (en **dB**) est alors définie comme suit :

$$L = 10 \log \frac{I}{I_0}$$

Où **I_0** le son le plus faible que l'on puisse entendre pour un signal de fréquence 1 kHz. Elle est appelée aussi intensité de référence.

$$I_0 = 10^{-12} \text{ W/m}^2$$

Pour que l'oreille perçoive un son dit audible, l'intensité sonore **I** doit être :

$$10^{-12} \text{ W/m}^2 \ll I \ll (1 \text{ à } 100 \text{ W/m}^2) \quad [21]$$

Où **10^{-12} W/m^2** est la limite de sensibilité de l'oreille et la borne supérieure de l'intensité sonore correspond à une destruction de l'oreille.

1.A.2.c. Timbre

Il correspond à la richesse d'un signal sonore d'un instrument ou de la voix d'une personne. Il est caractérisé par ses fréquences harmoniques, ses nombres et ses amplitudes. C'est pour cette raison qu'une même note ne produira pas le même son avec deux instruments différents. On dit qu'un son est riche lorsqu'il possède beaucoup d'harmoniques et pauvre lorsqu'il a moins d'harmoniques.

I.A.2.d. Phonème

Le phonème est la plus petite unité phonique distinctive, qui n'est définie sur une base acoustique, ni articulatoire, ni perceptuel, mais sur le plan fonctionnel. Ainsi, les phonèmes n'ont pas d'existence indépendante, c'est-à-dire, ils constituent un ensemble structuré dans lequel chaque élément est intentionnellement différent de tous les autres et la différence étant à chaque fois porteuse de sens.

Voici un exemple des phonèmes de la langue française. Dans la langue française, il existe 36 phonèmes.

CONSONNES	VOYELLES
Paie, baie, mais, fait, vais, ouais, taie, dais, nez, sait, huer, lait,	Lit, les, là, lin, lu, leu, leur, le, lent, loup, lot, lotte, long,

Tableau 2 Les phonèmes de la langue française

I.A.3. Prosodie

La prosodie est la façon de décrire ou de présenter formellement les éléments de l'expression orale à savoir les tons, les accents, l'intonation et la qualité. Ces éléments transmettent des informations sur la signification d'un énoncé. En d'autres termes, elle désigne les phonèmes liés à l'évolution dans le temps des paramètres de hauteur, d'intensité et de durée.

La perception de hauteur est essentiellement liée à la fréquence fondamentale qui correspond au niveau physiologique de la production et à la fréquence de vibration des cordes vocales.

La perception d'intensité est essentiellement liée à l'amplitude et à l'énergie du son, mais dépend aussi partiellement de sa durée.

La perception de durée correspond à son temps d'émission et sa durée acoustique. A noter que le terme « durée » est utilisé pour désigner à la fois le paramètre perceptif et le paramètre acoustique et le terme « longueur » comme synonyme de durée perçue est utile quand la distinction est importante.

I.A.4. L'information vocale [6] [14]

Le signal de la parole véhicule plusieurs types d'informations, tels que le fondamental, la prosodie, le timbre et les phonèmes. Par conséquent, ceci impose, aux systèmes de reconnaissance vocale, de n'extraire que l'information nécessaire à son application, les phonèmes pour les machines de dictée par exemple.

La parole est surtout contenue dans les deux premiers formants, mais l'information proprement dite provient des transitions formantiques.

En général, on considère que la plage de fréquences d'un signal de parole se situe dans la bande de 100Hz-5KHz (300Hz-3.4KHz pour la téléphonie). [14]

I.B. METHODE DE RECONNAISSANCE VOCALE

Les systèmes de reconnaissance vocale numériques (**Figure 5**) sont caractérisés par :

- le prétraitement qui comprend l'acquisition du signal de la parole et l'extraction des paramètres,
- l'apprentissage du vocabulaire et la comparaison aux références,
- le traitement des résultats.

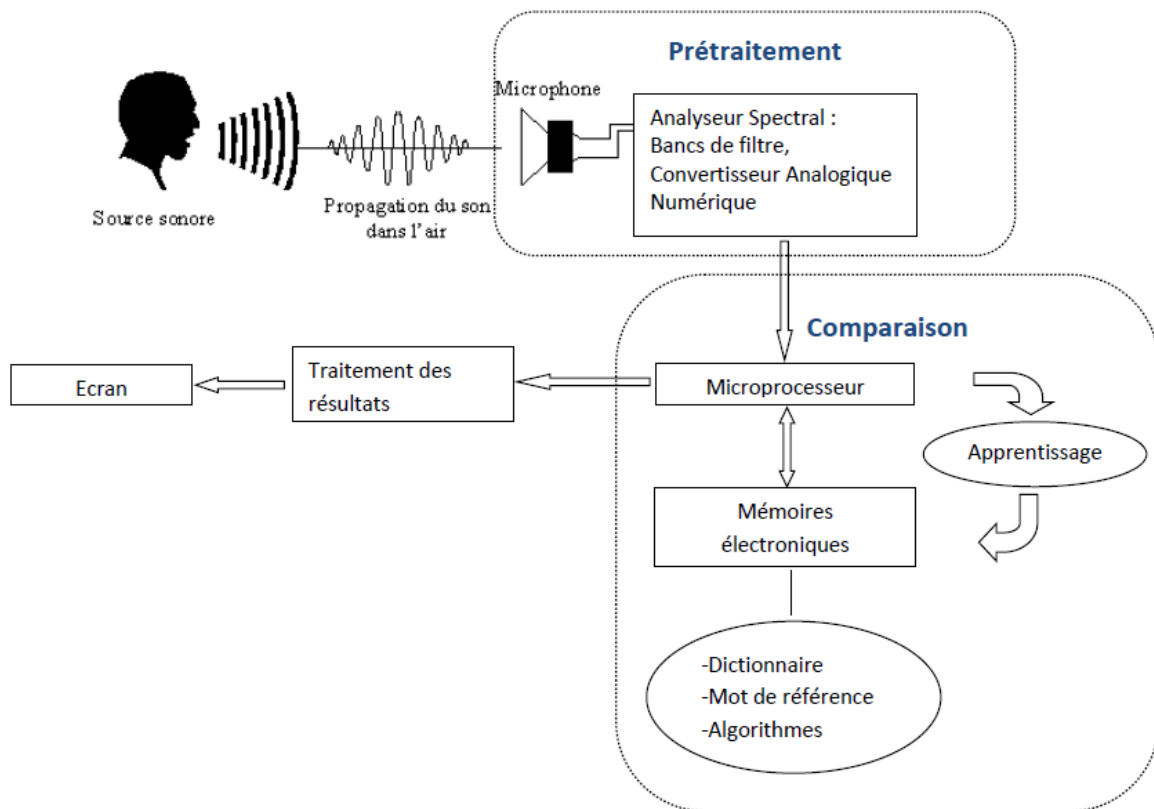


Figure 5 *Système de reconnaissance vocale*

La phase d'apprentissage pendant laquelle un ou plusieurs locuteurs prononcent une ou plusieurs fois chacun des mots de l'application prévue. Ces prononciations sont toutes prétraitées puis conservées telles quelles ou bien moyennées dans un dictionnaire de références en tant que “ images acoustiques ”.

Puis la phase de reconnaissance où le signal à reconnaître subit le même prétraitement que la phase précédente. Il est ensuite comparé aux références contenues dans le dictionnaire. Le calcul d'une « distance » et sa comparaison à un seuil permettent ou non de retenir la ou les références les plus proches.

Mais les différences de prononciation et les variations de débit d'élocution, parfois importantes et non linéaires imposent l'utilisation d'algorithmes de comparaison telle que la comparaison dynamique ou les chaînes de Markov.

I.B.1. Le prétraitement du signal

L'information portée par le signal de la parole est essentiellement contenue dans les formants. Il est donc nécessaire d'affranchir ces derniers de signaux indésirables tels que la source ou le « bruit » de numérisation (enveloppe spectrale en dents de scie).

Les systèmes de prétraitement permettent donc d'améliorer la représentation des formants par « lissage », et de fournir, à l'outil de comparaison, des vecteurs constitués de coefficients pertinents.

Les méthodes les plus courantes pour le traitement du signal de la parole sont les analyses spectrales réalisées par transformée de Fourier.

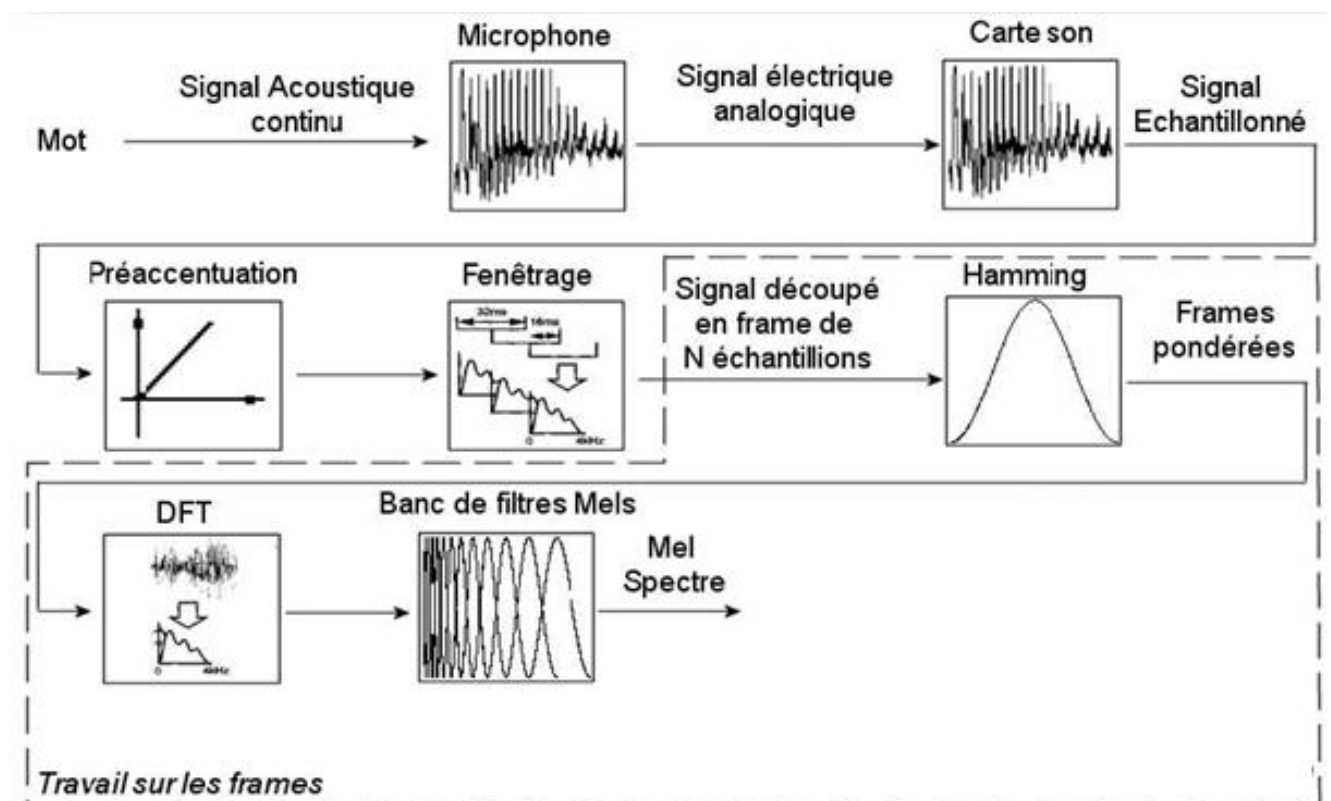


Figure 6 Schéma bloc du prétraitement [12]

1.B.1.a. La numérisation

La conversion d'un signal analogique sous forme numérique implique une double approximation. D'une part, dans l'espace des temps, le signal fonction du temps $x(t)$ est remplacé par ses valeurs $x(nT)$ à des instants multiples entiers d'une durée T ; c'est l'opération d'échantillonnage. D'autre part, dans l'espace des amplitudes, chaque valeur $x(nT)$ est approchée par un multiple entier d'une quantité élémentaire q ; c'est l'opération de quantification. La valeur approchée ainsi obtenue est ensuite associée à un nombre ; c'est le codage, ce terme étant souvent utilisé pour désigner l'ensemble, c'est-à-dire le passage de la valeur $x(nT)$ au nombre qui la représente.

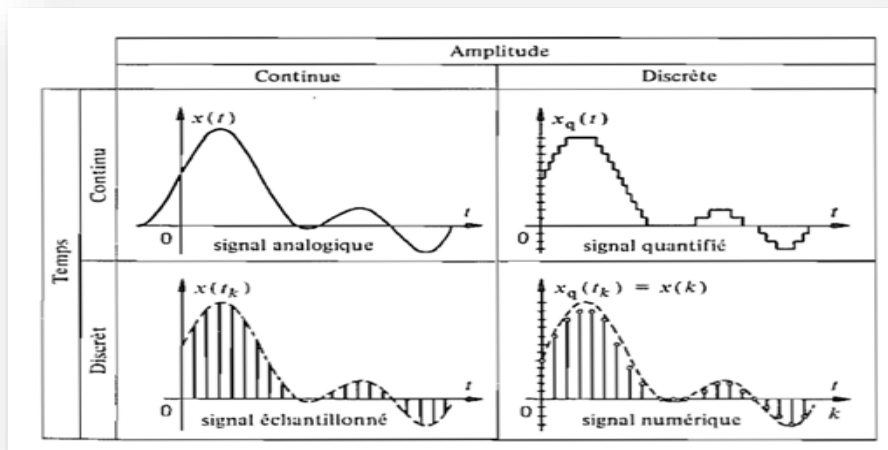


Figure 7 Classification morphologique des signaux

i. Opération d'échantillonnage [12]

1. Échantillonnage idéal

- Description temporelle**

L'opération d'échantillonnage consiste à prélever sur un signal analogique $x(t)$ dont l'évolution est continue dans le temps, des échantillons représentant l'amplitude aux instants de prélèvement.

Pour des raisons de simplification, les prélèvements sont réalisés régulièrement avec une période constante T_e appelée période d'échantillonnage. L'échantillonnage est

qualifié d'idéal (**Figure 8**) dès lors que l'on peut supposer ou approcher une prise instantanée des échantillons.

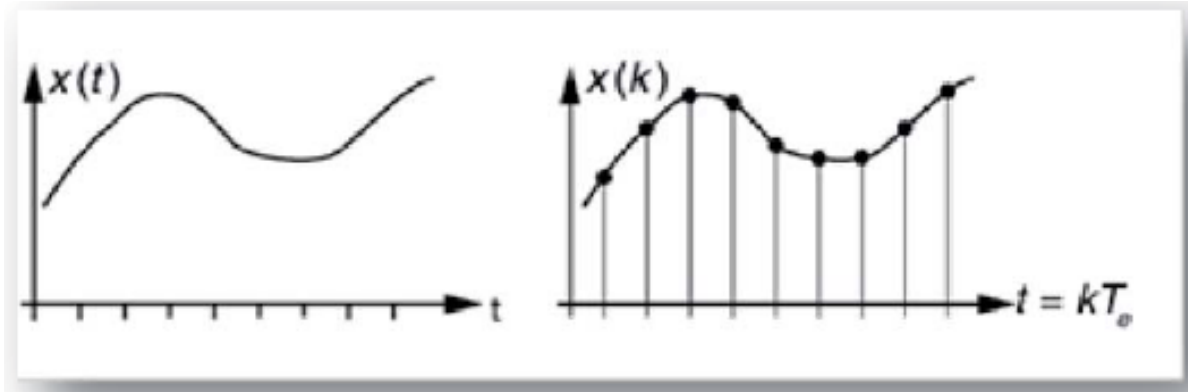


Figure 8 Echantillonnage idéal

Et l'expression du signal $x^*(t)$ après échantillonnage s'écrit :

$$x^*(t) = T_e x(t) \sum_{k=-\infty}^{k=+\infty} \delta(t - kT_e)$$

- **Description fréquentielle**

Soit $X^*(f)$ le transformé de Fourier appliqué au signal échantillonné $x^*(t)$ et on a :

$$X^*(f) = \sum_{k=-\infty}^{k=+\infty} X(f - kf_c) \text{ avec } f_c = \frac{1}{T_e}$$

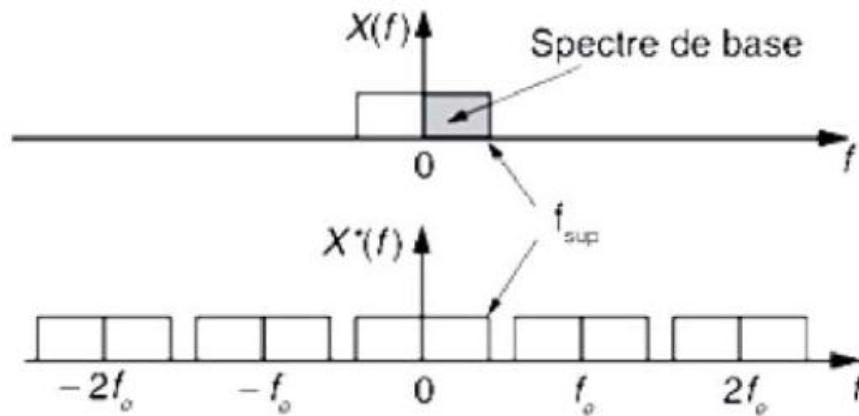


Figure 9 Composition spectral d'un signal

On remarque que le spectre du signal $x^*(t)$ est constitué d'une suite de répliques du spectre de $x(t)$ décalées avec une période T_e .

- **Théorème de l'échantillonnage**

En considérant un spectre initial $X(f)$ borné supérieurement par une limite f_{sup} , on peut espérer conserver toute l'information lorsque f est choisie telle que :

$$f_e > 2 f_{\text{sup}}$$

Le non-respect de la condition précédente conduit à un sous-échantillonnage qui engendre le repliement des motifs spectraux (**Figure 10**). En d'autres termes, il y a perte ou modification des informations originales.

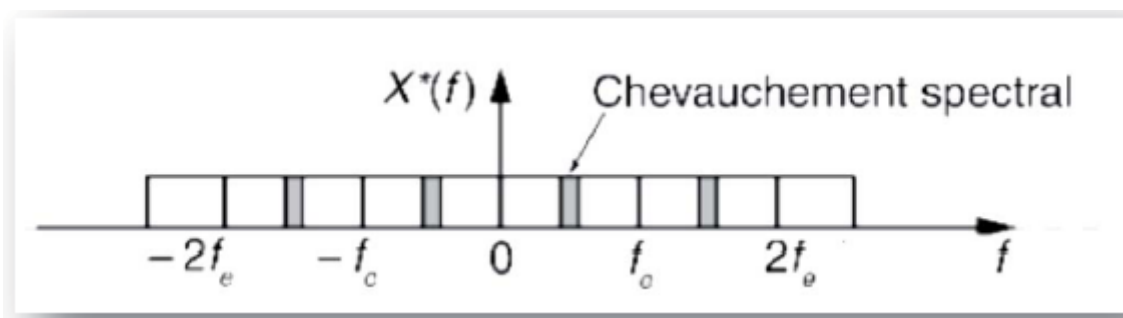


Figure 10 Aspect fréquentiel du sous – échantillonnage

2. Échantillonneur-bloqueur(E/B)

- Description

L'opération d'échantillonnage-blocage d'un signal $x(t)$ consiste à conserver la valeur échantillonnée entre deux prélèvements successifs, (**figure11**). Le signal issu d'un tel traitement prend une forme dite en marches d'escalier ; on le notera $x_n(t)$

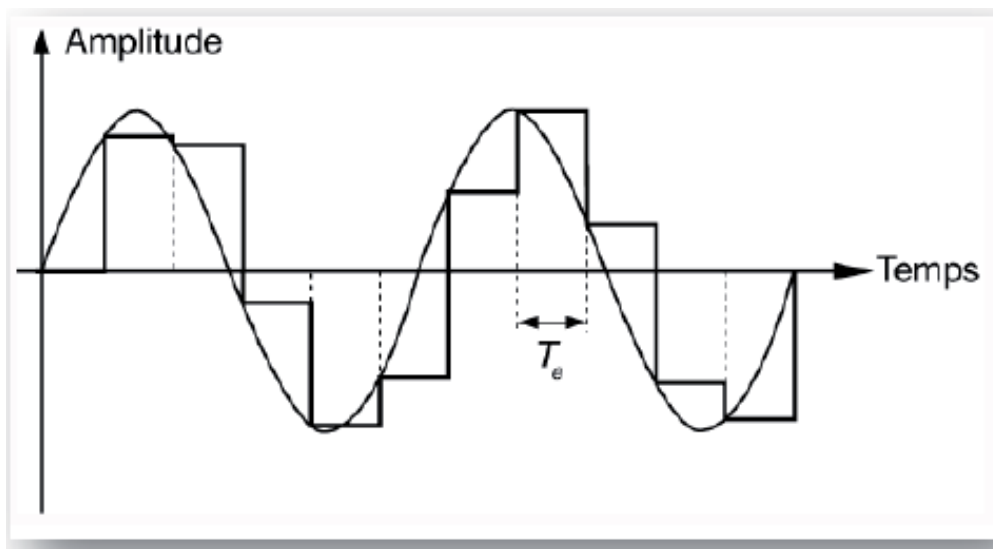


Figure 11 Echantillonneur-bloqueur d'un signal analogique

- Rôle d'Echantillonneur Bloqueur

L'échantillonneur-bloqueur est un pré conditionnement du signal analogique avant sa quantification par un Convertisseur Analogique-Numérique (CAN). Pendant la phase de conversion, un niveau de tension stable et constant doit être présent en entrée du CAN, afin d'éviter des erreurs durant la numérisation.

Considérons un signal analogique $x(t)$ à l'entrée d'un CAN. Soit t_c le temps nécessaire à la conversion et ΔX la variation de $x(t)$ pendant t_c .

$$\Delta X = t_c \frac{dx(t)}{dt}$$

Pour qu'il n'y ait pas erreur de conversion, il faut que ΔX soit inférieur au pas de quantification δ . Ce qui implique :

$$\tau_c \leq \frac{\delta}{\frac{dx(t)}{dt}}$$

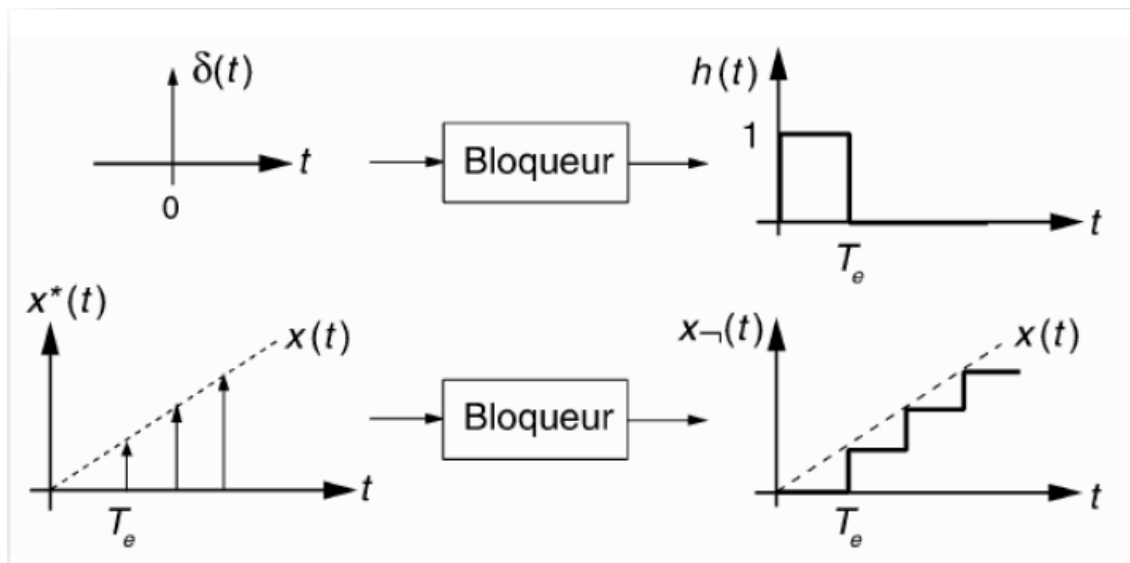


Figure 12 Description de la fonction blocage

ii. Opération de quantification

1. Quantification uniforme

La quantification est la seconde étape nécessaire à la numérisation des signaux. Succédant à l'échantillonneur-bloqueur, elle permet le traitement numérique ou la mémorisation du signal (**Figure 13**). Son rôle est d'affecter une valeur de résolution finie à un échantillon dont l'amplitude est en théorie infiniment précise si l'on fait abstraction du bruit de fond propre au signal.

Quantifier un échantillon, c'est arrondir sa valeur à celle de l'échelon le plus proche sur une grille de niveaux (**Figure 15**). Lorsque les échelons sont à pas constant, la quantification est uniforme.

L'étendue de la grille doit couvrir la gamme dynamique des signaux à convertir. Vis-à-vis des signaux, la quantification peut être modélisée par le schéma de la (**Figure14**) où :

- $Q(z)$ est la transformée en z d'un signal aléatoire ϵ_q appelé bruit de quantification.
- $X(z)$ et $X_q(z)$ sont respectivement les transformées du signal d'entrée et du signal quantifié.

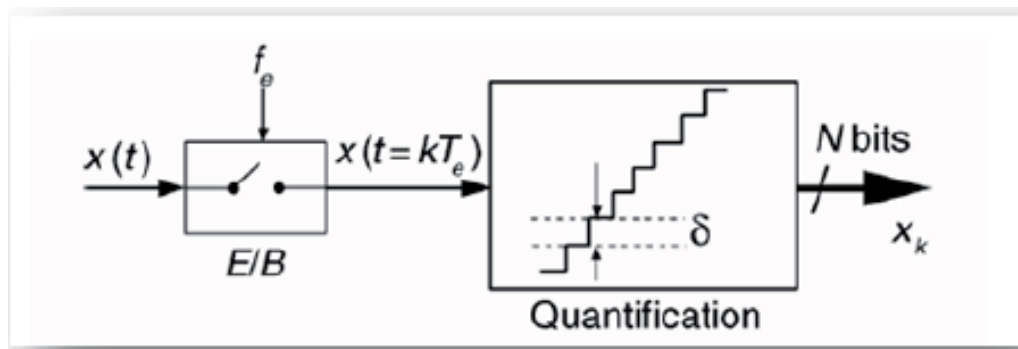


Figure 13 Chaîne de numération

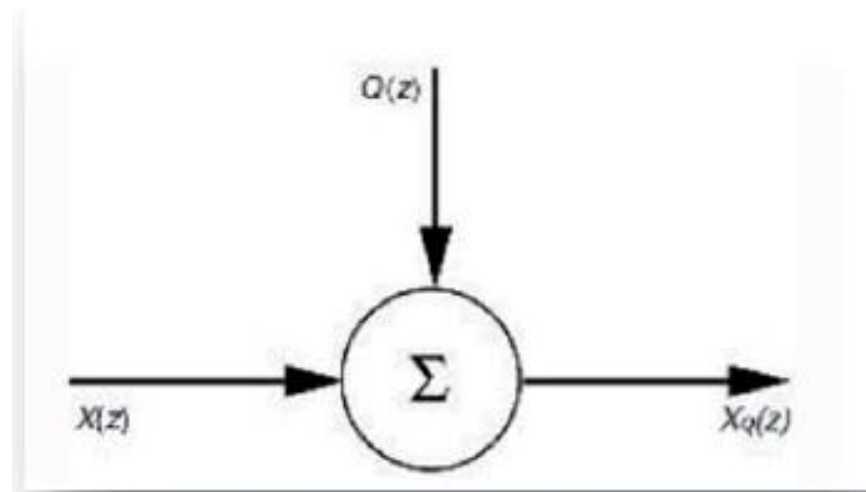


Figure 14 Modélisation de quantification

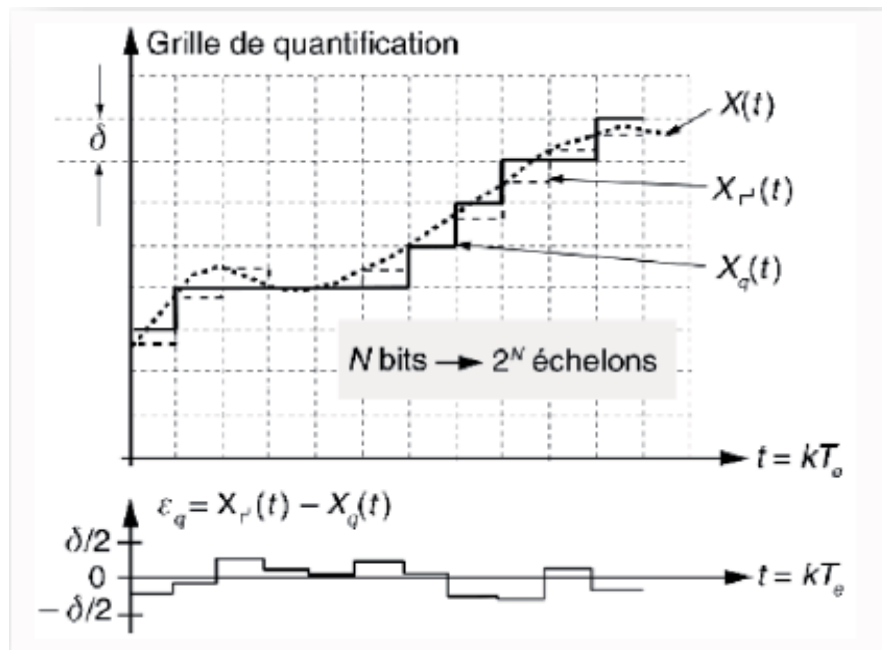


Figure 15 Quantification Uniforme et bruit de quantification

La quantification peut être par arrondi ou par défaut pour :

- la quantification par arrondi, toutes les valeurs du signal comprises entre

$$(n - \frac{1}{2})\delta < x < (n + \frac{1}{2})\delta \quad \text{est arrondi à} \quad x = n\delta$$

- la quantification par défaut ou par troncature, toutes les valeurs du signal comprises entre :

$$n\delta < x < (n + 1)\delta \quad \text{est arrondi à} \quad x = n\delta$$

La quantification est une opération non conservatrice car la précision originale ne peut être retrouvée après arrondi. Elle peut être vue comme la superposition d'une composante aléatoire sur l'amplitude de chaque échantillon original. Le bruit de quantification représente ainsi l'erreur liée à l'arrondi.

2. Bruit de quantification

- **Densité de probabilité**

Le bruit de quantification ε_q est une variable aléatoire. Lorsque le signal d'entrée a une amplitude suffisante (par exemple supérieure à 10 pas de quantification), on peut considérer que sa densité de probabilité $p(\varepsilon_q)$ est uniforme dans son domaine de variation, c'est-à-dire dans l'intervalle $\left[-\frac{\delta}{2}; \frac{\delta}{2}\right]$

$$\text{Sachant que } \int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} p(\varepsilon_q) d\varepsilon_q = 1 \quad \text{et on } p(\varepsilon_q) = \frac{1}{\delta}$$

- **Valeur moyenne du bruit de quantification**

Cette valeur est donnée par :

$$\int_{-\delta/2}^{\delta/2} p(\varepsilon_q) \varepsilon_q d\varepsilon_q = 0$$

- **Puissance du bruit de quantification**

La variance représente la puissance d'un signal aléatoire, elle est donnée par :

$$\sigma_q^2 = \int_{-\delta/2}^{\delta/2} p(\varepsilon_q) \varepsilon_q^2 d\varepsilon_q = \frac{\delta^2}{12}$$

iii. Codage

Le codage consiste à associer à un ensemble de valeurs discrètes un code composé d'éléments binaires. Le codage est la dernière étape et un outil très important à la numérisation d'un signal analogique.

Les codes les plus connus et les plus utilisés sont : code binaire naturel, code complément à 2, code DCB.

1. Code binaire naturel ou DCBN

Il est défini de la manière suivante :

$$N = \sum a_i 2^i$$

Avec $a_i \in \{0,1\}$ et $i \in \mathbb{N}$.

Ce code se prête parfaitement au traitement des opérations arithmétiques. Ses inconvénients sont les suivants :

- Il faut un grand nombre de bits pour exprimer un nombre dès que celui-ci est élevé.
- Ce code peut introduire des erreurs lors du codage des grandeurs variant de façon ordonnée. En effet entre deux mots successifs de ce code, plusieurs bits pourront être amenés à changer simultanément

2. Code DCB

Chaque élément d'un nombre décimal (chiffre décimal) est représenté par son équivalent binaire naturel à 4 bits.

Exemple : 1 9 8 7 \Rightarrow 0001 1001 1000 0111

Caractéristiques :

- C'est un code pondéré.
- Il conserve les avantages du système décimal et du code binaire pur.
- Les mots codes sont plus longs qu'en code DCBN.

On considère à partir de maintenant que le signal a été numérisé (ou échantillonné). Il est représenté à partir de maintenant par une famille $(x_n)_{n \in [1,M]}$ où M est le nombre total d'échantillons dans le signal.

I.B.1.b. Analyse spectrale des signaux numériques

L'objectif de l'analyse spectrale est la détermination des composantes fréquentielles d'un signal analogique ou numérique. Sur le plan théorique, l'analyse spectrale est largement inspirée de la transformation analytique de Fourier.

i. Transformation de Fourier (TF)

1. Définition

La Transformation de Fourier est une extension de la décomposition en série de Fourier, mais pour des signaux quelconques. Intuitivement on peut considérer un signal non périodique comme un signal dont la période $T \rightarrow +\infty$. Ainsi la somme discrète et le facteur $\frac{1}{T}$ intervenant dans la décomposition en série de Fourier deviennent respectivement une intégrale, et une petite variation de fréquence df . On définit la TF notée $X(f)$ du signal $x(t)$ par :

$$\begin{aligned} TF \{x(t)\} = X(f) &= \int_{-\infty}^{+\infty} x(t) \exp(-j\omega t) dt, \text{ avec } \omega = 2\pi f \\ TF^{-1} \{X(f)\} = x(t) &= \int_{-\infty}^{+\infty} X(f) \exp(+j\omega t) df \end{aligned} \quad (*)$$

Et on obtient :

- Spectre d'amplitude = $|X(f)|$
- Spectre(ou densité spectrale) de puissance = $|X(f)|^2$
- Spectre de phase = $Arctg \left(\frac{\text{Im} \{X(f)\}}{\text{Re} \{X(f)\}} \right) = \varphi(f)$

2. Conditions d'existence de la TF

Une fonction $x(t)$ admet pour Transformée de Fourier $X(f)$ si :

i. $x(t)$ bornée

ii. $\int_{-\infty}^{+\infty} x(t) dt$ existe

(**)

iii. Les discontinuités de $x(t)$ sont en nombre limité

Une grande partie des signaux étudiés répondent à ces conditions. Ceci est dû, en partie, au fait qu'ils sont observés sur une durée finie.

ii. Transformation de Fourier Numérique(TFN)

La Transformation de Fourier Numérique s'introduit quand il s'agit de calculer la Transformée de Fourier d'une fonction à l'aide d'un calculateur numérique. En effet un tel opérateur ne peut traiter que des nombres et de plus en quantité limitée par la taille de sa mémoire.

Soit une suite numérique de N valeurs x_k pour k variant de 0 à $N - 1$.

On peut toujours considérer que ces valeurs sont issues de l'échantillonnage à la cadence f_e

du signal $x(t)$ c'est-à-dire que $x_k = x(t = \frac{k}{f_e})$. La version numérique de l'équation (*)

devient :

$$X(f) = \sum_{k=0}^{N-1} x_k \exp(-j2\pi k \frac{f}{f_e}) \quad (\text{TFN})$$

$$x_k = \frac{1}{f_e} \int_{-f_e/2}^{f_e/2} X(f) \exp(j2\pi k \frac{f}{f_e}) df, \quad k \text{ variant de } 0 \text{ à } N - 1 \quad (\text{TFN inverse})$$

iii. Transformation de Fourier Discrète (TFD)

1. Discrétisation de l'intervalle fréquentiel $\left[-\frac{f_e}{2}, \frac{f_e}{2}\right]$

La fréquence f est assimilée à une variable continue dans la TFN ce qui ne rend pas aisé son calcul de manière informatisée. C'est pourquoi on choisit couramment de ne calculer que sur N fréquences discrètes réparties régulièrement dans l'intervalle fréquentiel de base.

On pose le pas fréquentiel : $\Delta F = \frac{f_e}{N}$

Et les fréquences discrètes f_n sont données par :

$$f_n = n\Delta f = n \frac{f_e}{N} \text{ ou } \frac{f_n}{f_e} = \frac{n}{N}$$

avec n un nombre entier déterminé dans la **(Figure 16)**.

On obtient directement la relation suivante qui est appelée Transformation de Fourier Discrète :

$$X_n = X(f_n) = \sum_{k=0}^{N-1} x_k \exp(-j2\pi k \frac{n}{N}) = \sum_{k=0}^{N-1} x_k W_N^{-nk} \quad (\text{TFD})$$

$$x_k = \frac{1}{N} \sum_{n=\frac{N}{2} \text{ (ou } \frac{N-1}{2})}^{\frac{N}{2}-1 \text{ (ou } \frac{N-1}{2})} X_n W_N^{nk}, \text{ k variant de 0 à N-1} \quad (\text{TFD inverse})$$

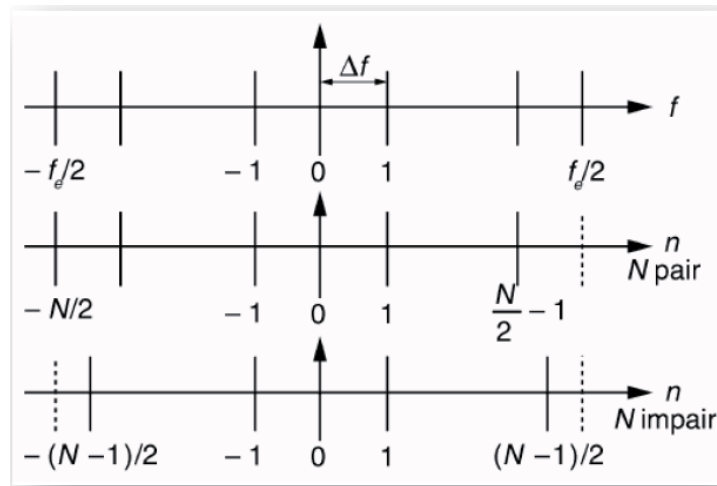


Figure 16 Répartition des fréquences discrètes

I.B.1.c. Parole en tant que signal quasi-stationnaire

La parole est un son très compliqué dans son ensemble, mais si on considère des intervalles de temps très réduits (environ 20 ms), le signal vocal est alors presque stationnaire. C'est pour cela que l'on considère le signal vocal comme un signal quasi stationnaire, en première approximation. On pourra ainsi utiliser des fonctions de fenêtrage pour améliorer les résultats.

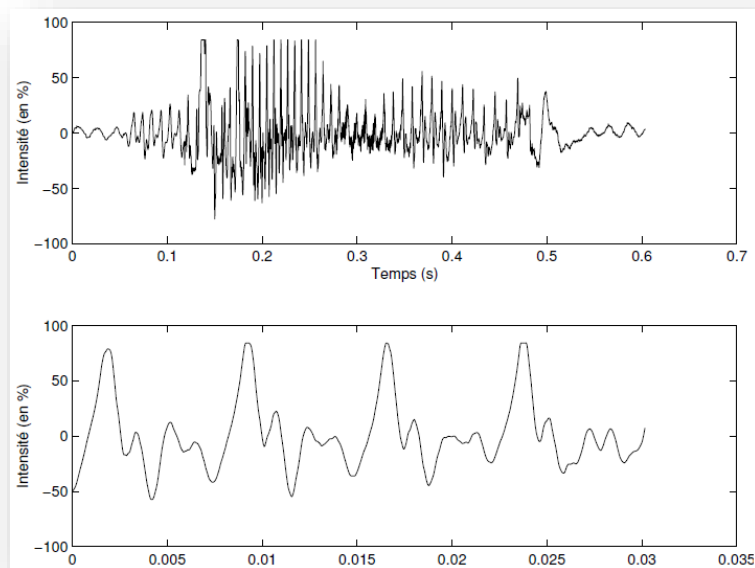


Figure 17 Représentation temporelle du mot "physique" (en haut) ainsi qu'un extrait de 30 ms pendant la prononciation de [Ph] (en bas)

On remarque bien que le signal peut être considéré comme périodique.

1.B.1.d. Fenêtrage

Le fenêtrage est une manière nécessaire pour obtenir une bonne qualité de spectre, il permettra d'obtenir une bonne information fréquentielle et une bonne analyse dans le domaine temporel.

Avant tout traitement, on procède à la préaccentuation du signal pour pallier le fait que les hautes fréquences seront moins puissantes que les basses fréquences :

$$h_n = 1 - \alpha z_n^{-1} \text{ .étant } \alpha \text{ le facteur de préaccentuation.}$$

Puis on effectue le fenêtrage du signal. Le signal échantillonné est découpé en tranches de 20 à 30 ms appelées « frames », qui ont la particularité de se recouvrir par moitié pour atténuer les discontinuités qui pourraient apparaître. On utilise typiquement un frame de $N = 256$ ou 512 échantillons, ou un nombre qui est une puissance de 2. Cela vient du fait que l'algorithme FFT, que l'on utilise ensuite, est plus rapide pour ces nombres.

A partir de la famille $(x_n)_{n \in [1, M]}$ on définit un ensemble de I familles $(x'_{i,n})_{n \in [1, N]}$ avec N le nombre d'échantillons par frame, et $i \in [1, I]$, I le nombre de frames.

Puis une fonction de fenêtrage est appliquée à chaque frame (généralement Hamming ou de type Triangulaire). Ceci pour minimiser les discontinuités au début et à la fin de chaque frame.

Ainsi, si $(w_n)_{n \in [1, N]}$ (N le nombre d'échantillons par frame) est la famille contenant les coefficients de pondération de Hamming et $(x'_{i,n})_{n \in [1, N]}$ est la $i^{\text{ème}}$ frame (aussi composée de N échantillons), le contenu $y_{i,n}$ de la $i^{\text{ème}}$ frame pondérée sera :

$$\forall n \in [1, N], y_{i,n} = x'_{i,n} \times w_n$$

Formule de Hamming :

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

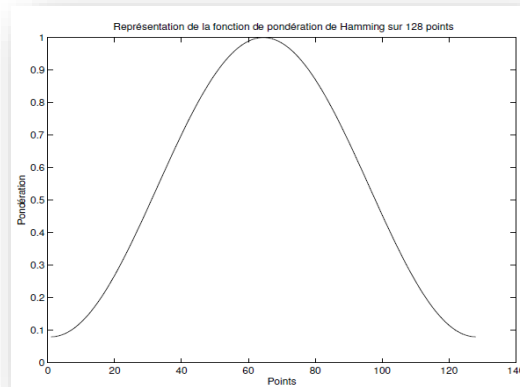


Figure 18 Courbe de la fonction de pondération de Hamming sur 128 points [16]

Les avantages du fenêtrage par Hamming sont l'amélioration de la netteté des harmoniques et supprime les discontinuités sur les bords (effet de bord).

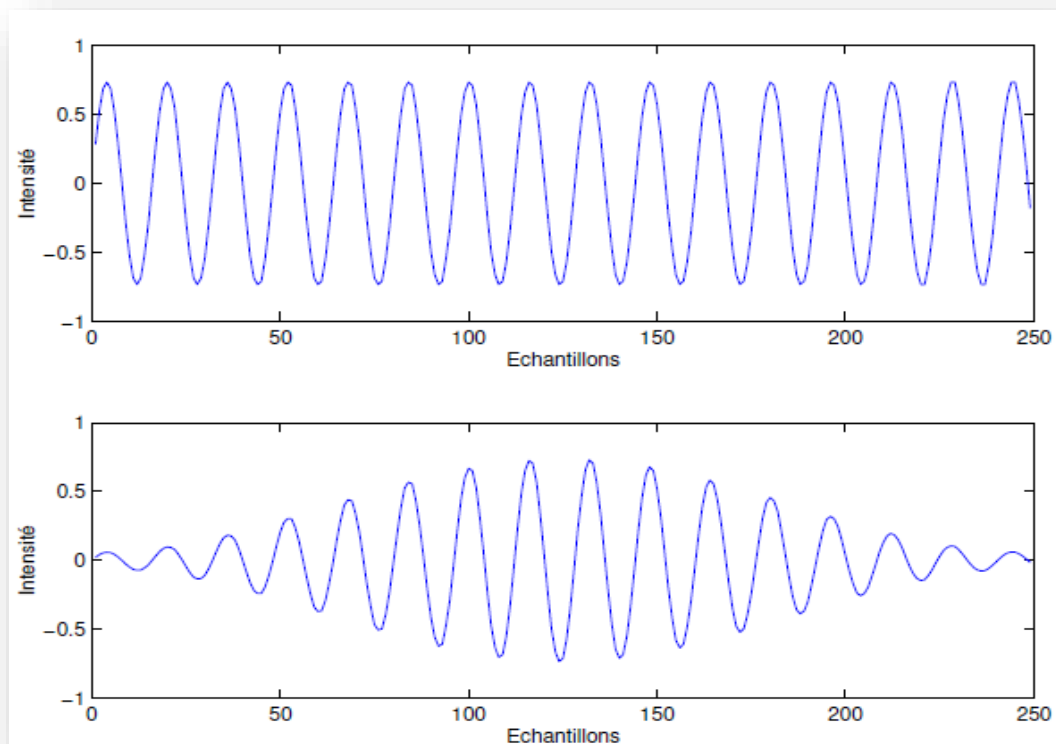


Figure 19 Représentation d'un signal sinusoïdal non-pondéré puis pondéré par la fonction de Hamming

1.B.1.e. Discret Fourier Transform(DFT)

L'étape suivante consiste à prendre la Transformée de Fourier de chaque frame fenêtrée (on utilise ici la Transformée de Fourier Discrète c'est-à-dire la version pour signal discret de la Transformée de Fourier). On convertit ainsi chaque frame du domaine temporel vers le domaine fréquentiel.

La DFT sur le i ème frame est définie ainsi :

$$Y_{i,n} = \sum_{k=1}^N y_{i,k} \exp\left(\frac{-2\pi j(n-1)(k-1)}{N}\right), \text{ avec } n \in [1, N] \text{ et } k = [0, 1, 2, \dots, N-1], N \text{ le}$$

nombre d'échantillons temporels.

1.B.1.f. Banc de filtres Mels

Parce que l'étendue des fréquences présentes dans le spectre est encore très large, donc beaucoup de données à traiter, on a recours au banc de filtres espacés selon l'échelle de Mels (**Figure 20**). On relie ainsi le système de reconnaissance au fonctionnement de l'oreille humaine.

Tout d'abord, la bande de fréquences utilisée se situe entre 0 et $\frac{F_e}{2}$ Hz. Cette bande va être divisée en k filtres triangulaires équidistants dans le domaine de fréquence Mel avec un recouvrement de 50%.

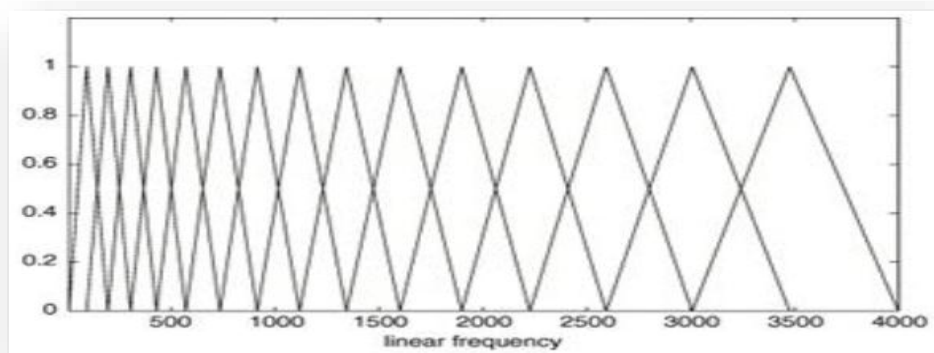


Figure 20 Banc de filtres espacés selon l'échelle de Mel (ici $F_e=8000$ Hz) [11]

La suite explique les différentes étapes et formules utilisées afin d'obtenir le banc de filtres. Tout d'abord, il faut savoir que l'échelle de fréquence Mel est espacée linéairement jusqu'à 1000 Hz et ensuite elle est espacée logarithmiquement au-dessus de 1000 Hz (**Tableau 3**). La transformation de la fréquence en fréquence de Mel s'effectue de la façon suivante :

$$Mel(f) = 25295 \times \log_{10}\left(1 + \frac{f}{700}\right)$$

Mel	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000	2100
Hz	510	630	770	920	1080	1270	1480	1720	2000	2320	2700	3150	3700	4400	5300	7700	9500

Tableau 3 Echelle de fréquence Mel

Et l'opération inverse :

$$Mel^{-1}(f_{mel}) = 700 \times \left(10^{\left(\frac{f_{mel}}{25295}\right)} - 1\right)$$

Pour connaître l'intervalle de fréquence Mel entre chaque filtre, on utilise la constante :

$$Mel_{intervalle} = \frac{Mel(f_{max}) - Mel(f_{min})}{K + 1} \quad \text{Où,}$$

$Mel(f_{max})$ = fréquence maximale sur l'échelle de Mel calculée à partir de f_{max}

(souvent $f_{max} = \frac{F_e}{2}$ Hz)

$Mel(f_{min})$ = fréquence minimale sur l'échelle de Mel calculée à partir de f_{min}

(Souvent $f_{min} = 0$ Hz)

K = nombre de filtres.

La fréquence centrale (en Hz) de chaque filtre k est obtenue par :

$$f_c(k) = \text{Mel}^{-1}(k \times \text{Mel}_{\text{intervalle}}) \text{ avec } 1 \leq k \leq K$$

En termes d'indices FFT pour la fréquence centrale des filtres, on a :

$$\text{index}_c(k) = \text{round}\left(\frac{f_c(k)}{F_e} \times N_{\text{FFT}}\right) \text{ avec } 1 \leq k \leq K$$

$\text{round}(\dots)$ = Arrondi à l'entier le plus proche

N_{FFT} = taille de la fenêtre FFT

Ensuite, on va pondérer le banc de filtres Mel en fonction d'où se trouve l'indice n de fréquence :

$$M_{n,k} = \left\{ \begin{array}{ll} 0 & , \text{pour } n < \text{index}_c(k-1) \\ \frac{n - \text{index}_c(k-1)}{\text{index}_c(k) - \text{index}_c(k-1)} & , \text{pour } \text{index}_c(k-1) \leq n < \text{index}_c(k) \\ \frac{n - \text{index}_c(k+1)}{\text{index}_c(k) - \text{index}_c(k+1)} & , \text{pour } \text{index}_c(k) \leq n < \text{index}_c(k+1) \\ 0 & , \text{pour } n > \text{index}_c(k+1) \end{array} \right\}$$

Et on obtient la représentation graphique de $M_{n,k}$:

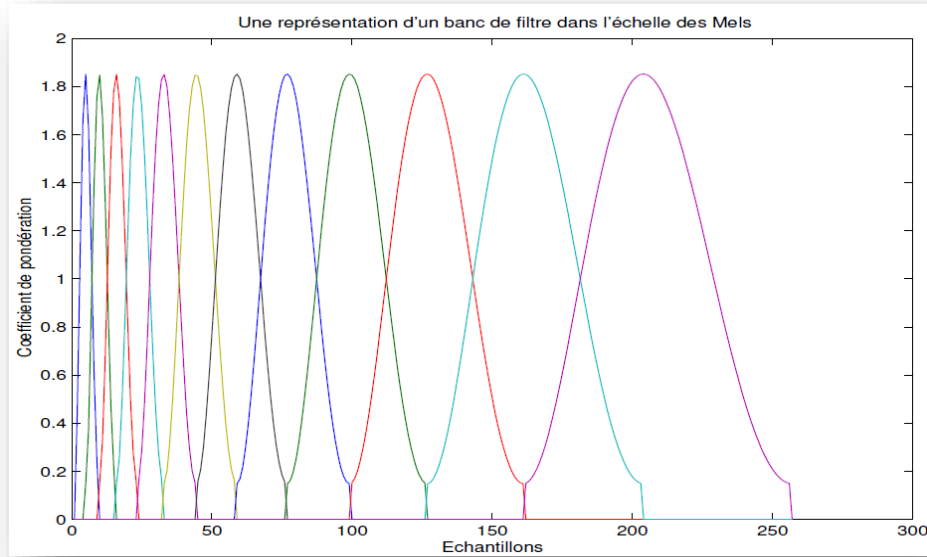


Figure 21 Représentation de la matrice $N \times K$ des coefficients du banc de filtre (N : nombre d'échantillons par frame, K nombre de filtres souhaités) [11]

Chaque filtre va donner un coefficient :

$$S_{i,k} = \sum_{n=0}^{\frac{N}{2}} |Y_{i,n}| M_{n,k} \quad k = 1 \dots K$$

$Y_{i,n}$ = le $n^{\text{ème}}$ $\in [1, N]$ coefficient de la transformée de la $i^{\text{ème}}$ $\in [1, I]$ frame.

$M_{n,k}$ = le $n^{\text{ème}}$ $\in [1, N]$ coefficient du $k^{\text{ème}}$ $\in [1, K]$ filtre.

$S_{i,k}$ = la matrice de sortie du $k^{\text{ème}}$ filtre pour la $i^{\text{ème}}$ frame.

On a à cette étape, ce que l'on appelle un Spectre Mel (Mel Spectrum).

I.B.2. La comparaison Dynamique ou DTW

Lorsqu'un locuteur prononce deux fois un même mot, les spectrogrammes ne seront jamais exactement les mêmes. Il y aura des différences non linéaires dans le temps (rythme) qui nécessite « d'aligner les axes temporels ».

Soient A et B deux images acoustiques de longueur I et J respectivement. La « distance » entre l'évènement $i \in [1, I]$ de A et l'évènement de B se calcule avec une simple distance euclidienne :

$$d(i, j) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad \text{avec } i = \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_N \end{pmatrix} \quad \text{et } j = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{pmatrix}$$

Cela suppose que l'on considère la même plage de fréquence pour les 2 signaux (entre 0 et N).

On crée donc un chemin $\{C(K) = (n(k), m(k)), \forall k \in [1, K]\}$. Il est nécessaire que les fonctions $n(k)$ et $m(k)$ soient croissantes et doivent correspondre à certaines contraintes : les seuls chemins valides arrivant au point (i, j) sont ceux provenant des points

$(i-1, j)$, $(i, j-1)$ et $(i-1, j-1)$. De plus on prendra K tel que $C(K) = (I, J)$. On pose $C(1) = (1, 1)$

La méthode consiste à choisir le chemin qui passe par les distances $d(i, j)$ les plus petites, de sorte que la distance cumulée le long de ce chemin soit la plus petite possible.

On définit $g(i, j)$ la distance cumulée au point (i, j) comme :

$$g(i, j) = \min \begin{cases} g(i-1, j) + d(i, j) \\ g(i-1, j-1) + 2 \cdot d(i, j) \\ g(i, j-1) + d(i, j) \end{cases}$$

On remplit ensuite la matrice $I \times J$ (le plan du chemin) avec $i^{\text{ème}}$ et $j^{\text{ème}}$ colonnes le résultat de $g(i, j)$. Enfin, on définit la distance normalisée entre les deux spectres :

$$G = \frac{g(I,J)}{I+J}$$

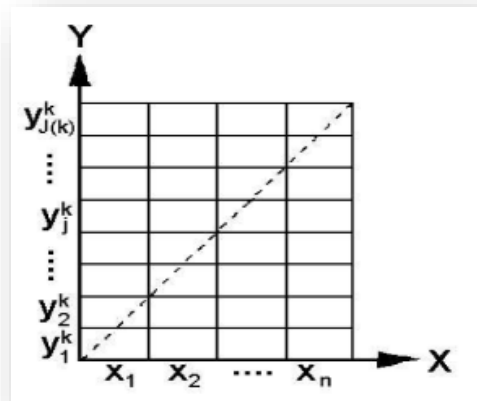


Figure 22 *Chemin idéal entre 2 spectres* [6]

Chapitre II. Réalisation du didacticiel nommé « PC Helper »

Le but de cette réalisation est de donner quelques exemples d'implémentations informatiques des principes physiques et mathématiques qu'on a vu antérieurement. A ce sujet, On a développé une application Android nommée « PC Helper » pour montrer la simulation sous Tablette.

II.A. PROGRAMMATION

Android est un système d'exploitation mobile open source. Le système a d'abord conçu pour les smartphones et tablette tactiles, puis s'est diversifié dans les objets connectés et ordinateurs comme Android TV, Android Auto Android Wear. En 2015, Android est le système d'exploitation le plus utilisé dans le monde avec plus de 80% de parts de marché dans les smartphones et tablette.

Le développement d'applications pour Android s'effectue avec un ordinateur sous Mac OS, Windows, ou Linux en utilisant le JDK de la plate-forme JAVA et des outils pour Android. Ces outils sont mis à disposition sous la forme d'un plugin pour l'environnement de développement Eclipse.

Grâce à cette nouvelle technologie plus performante, l'application reconnaissance vocale est développée sous Android et voici quelques spécifications techniques de l'application « PC Helper »

Terminal mobile	Tablette
Android version minimale	3.2
Orientation	Landscape

Tableau 4 Spécifications techniques du logiciel « PC Helper »

II.B. LES INTERFACES DE « PC Helper »

II.B.1. Ecran de démarrage



Figure 23 Ecran de démarrage

Il affiche juste après le lancement de l'application dans un délai de 5 secondes. Il permet de voir tout de suite le nom de l'application et son réalisateur.

II.B.2. Ecran de connexion

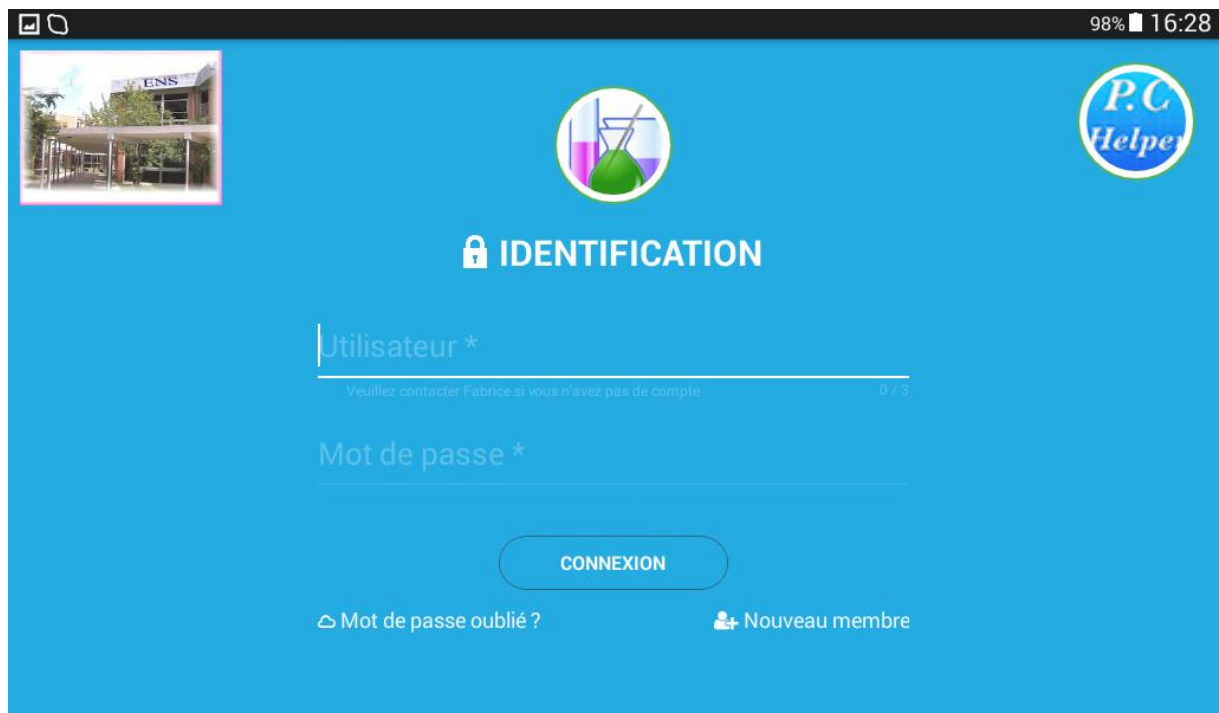


Figure 24 Ecran de connexion

Il permet de saisir le nom d'utilisateur et son mot de passe avec un contrôle de haut niveau. En tapant sur le bouton « **connexion** », l'application vérifie si le compte utilisateur est autorisé ou non. Si ce compte est validé, il peut continuer à utiliser l'application sinon il doit demander au réalisateur de cette application d'avoir une autorisation.

II.B.3. Ecran d'accueil

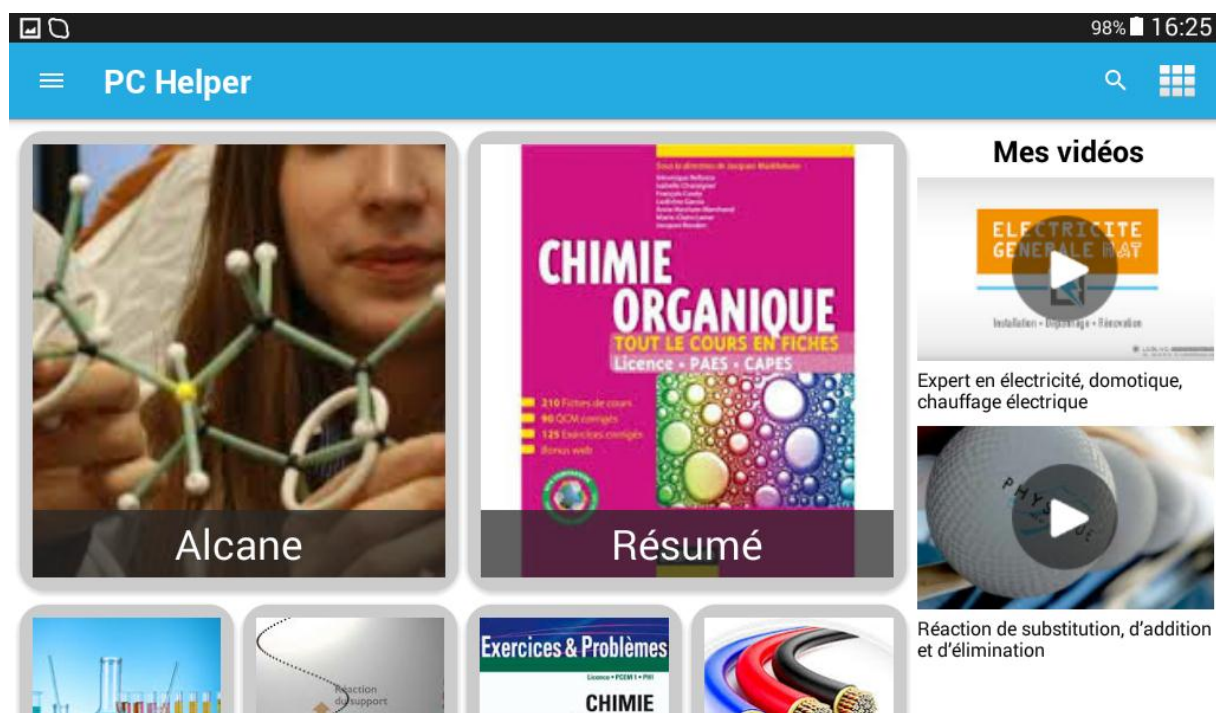


Figure 25 Ecran d'accueil

C'est l'écran principal de l'application. La rubrique « **Mes vidéos** » permet de visualiser les vidéos dont les titres sont indiqués au-dessous de l'image dans une vue plus dynamique. En cliquant sur l'image intitulée « **Mécanique** », par exemple, l'application affiche les informations plus détaillées de ce sujet.

II.B.4. Ecran détail

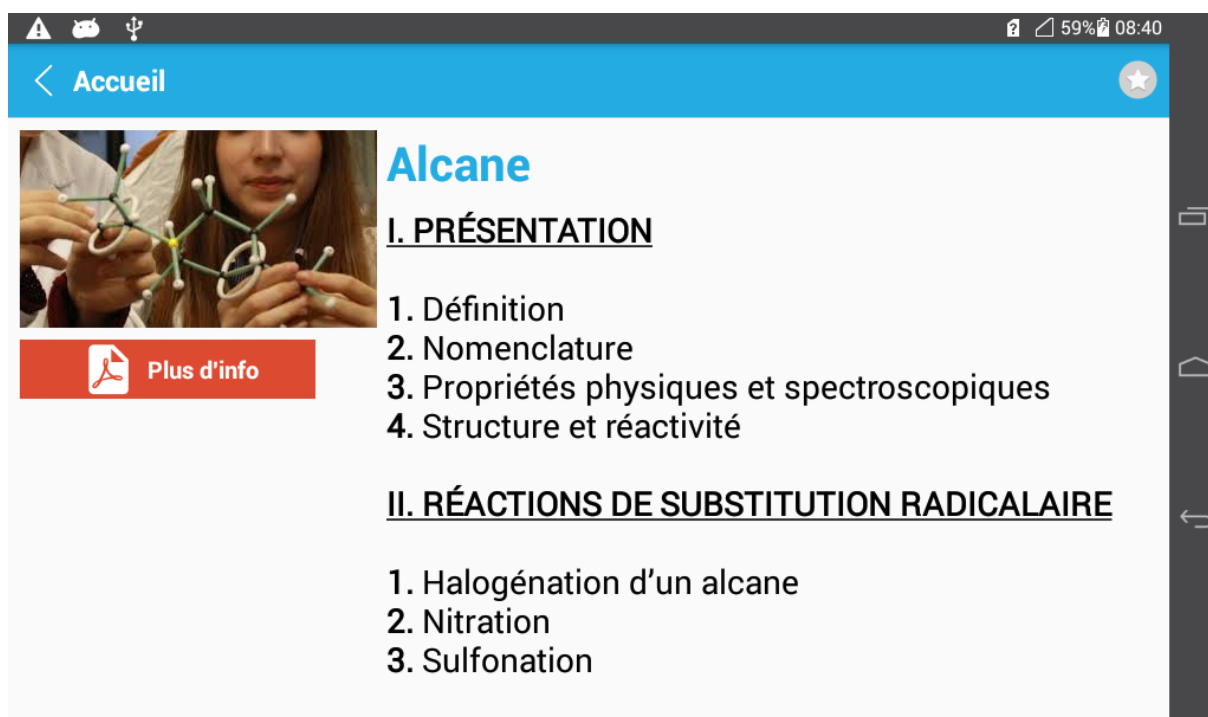


Figure 26 Ecran détail

Il illustre les informations plus détaillées des items listés à l'écran d'accueil.

Le bouton « **étoile** » permet d'ajouter parmi la liste des favoris.

Le bouton « **Plus d'info** » permet d'afficher le document approprié.

II.B.5. Ecran Vidéo

59% 08:40

< Réaction de substitution, d'addition et d'élimination

LES GRANDS TYPES DE RÉACTIONS EN CHIMIE ORGANIQUE -

- Substitution.
 $\text{CH}_3\text{-Cl} + \text{H}_2\text{O} \rightarrow \begin{array}{c} \text{CH}_3 \\ | \\ \text{C}-\text{OH} \\ | \\ \text{CH}_3 \end{array} + \text{H}^+ + \text{Cl}^-$
- Addition.
 $\text{CH}_2=\text{CH}_2 + \text{Br}_2 \rightarrow \begin{array}{cc} \text{Br} & \text{Br} \\ | & | \\ \text{CH}_2 & - & \text{CH}_2 \\ / & \backslash \\ \text{H} & \text{H} \end{array}$
- Élimination.
 $\begin{array}{c} \text{CH}_3 \\ | \\ \text{CH}_3-\text{C}-\text{CH}-\text{CH}_3 \\ | \quad | \\ \text{OH} \quad \text{H} \end{array} \rightarrow \begin{array}{c} \text{CH}_3 \\ | \\ \text{CH}_3-\text{C}=\text{CH}-\text{CH}_3 \end{array} + \text{H}_2\text{O}$




Figure 27 Ecran vidéo

II.B.6. Menu

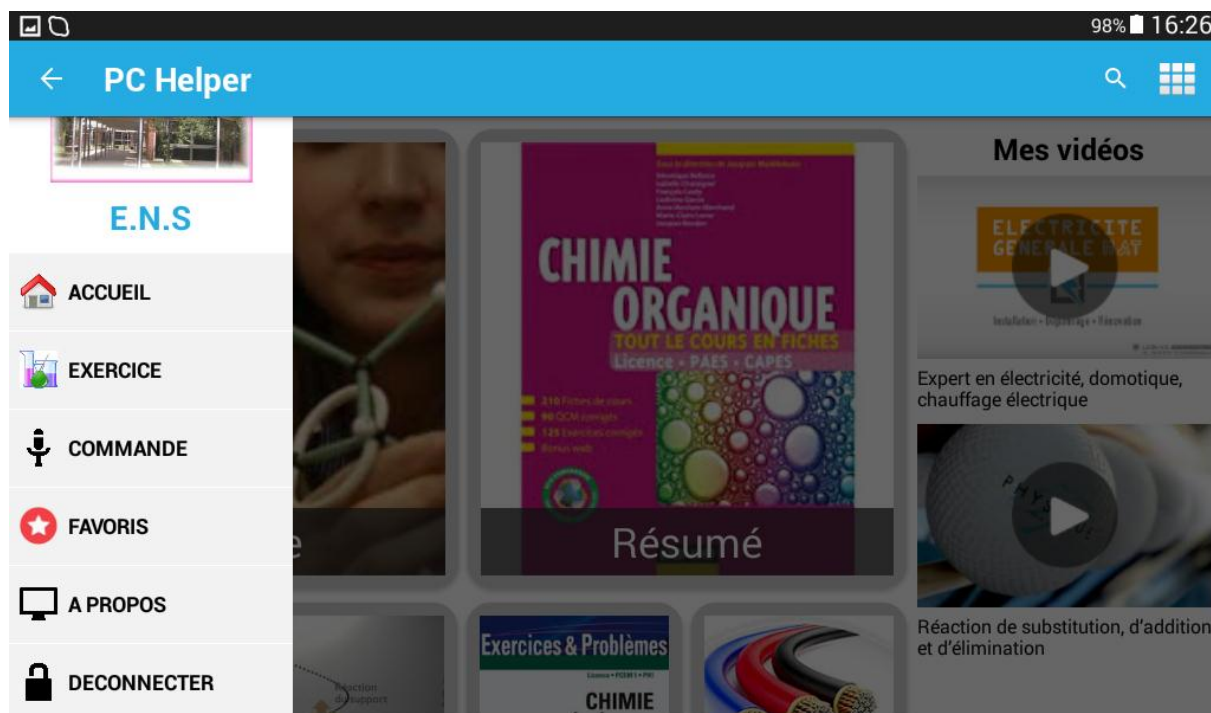


Figure 28 Ecran menu

Il permet de naviguer entre les différents écrans qui constituent l'application.

II.B.7. Ecran EXERCICE

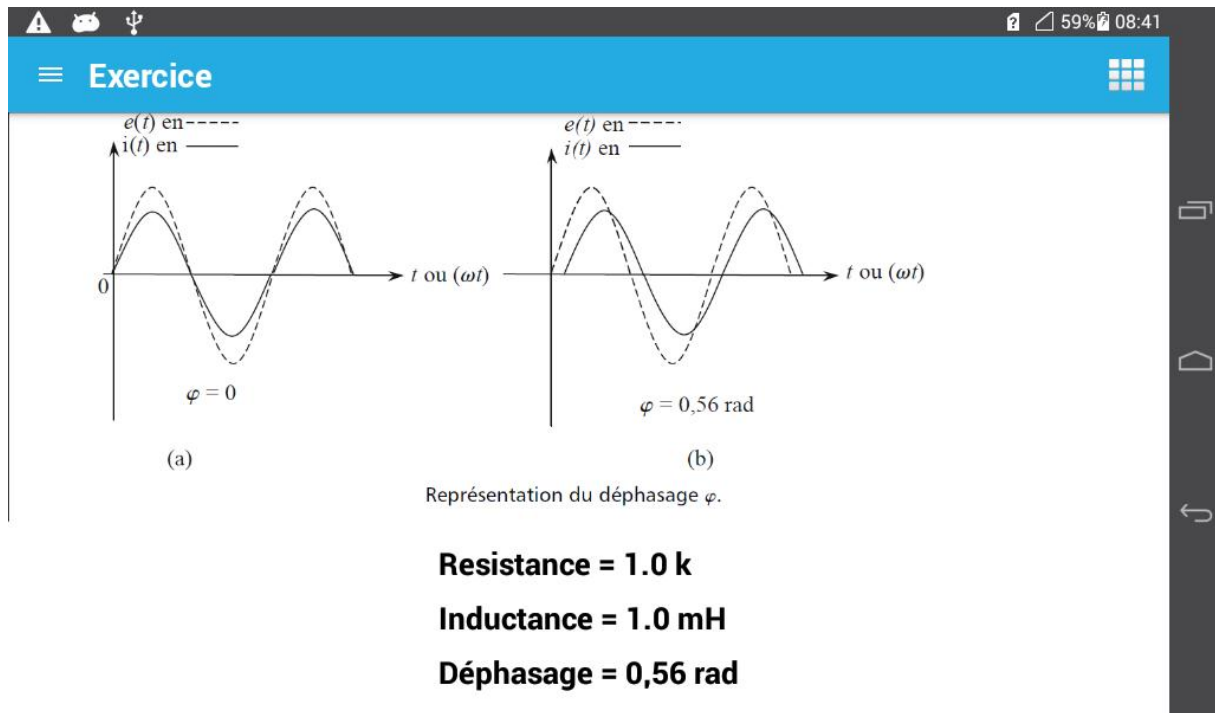


Figure 29 Ecran exercice

Il illustre un exercice type sur l'étude du déphasage entre le courant et la tension dans un régime du courant alternatif. En cliquant sur la valeur de la **résistance** ou de l'**inductance**, on peut insérer via la commande vocale la valeur de la résistance ou de l'inductance voulue. L'application calcule automatiquement le déphasage correspondant à cette valeur dès que la commande est terminée.

II.B.8. Ecran COMMANDE

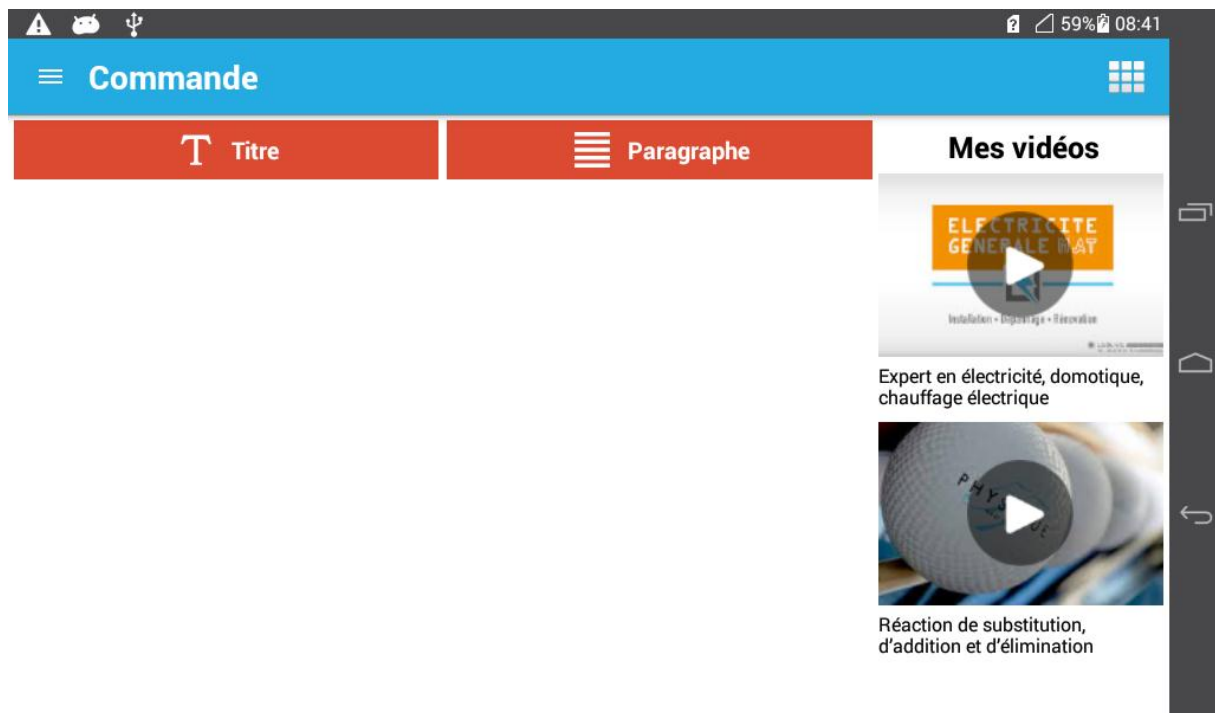


Figure 30 Ecran commande

Il permet à l'utilisateur de transformer sa commande vocale en texte sous forme de titre ou de paragraphe selon son choix. L'utilisateur utilise donc sa voix pour prendre de note, à titre d'exemple, au lieu d'écrire par sa main.

II.B.9. Ecran Favoris

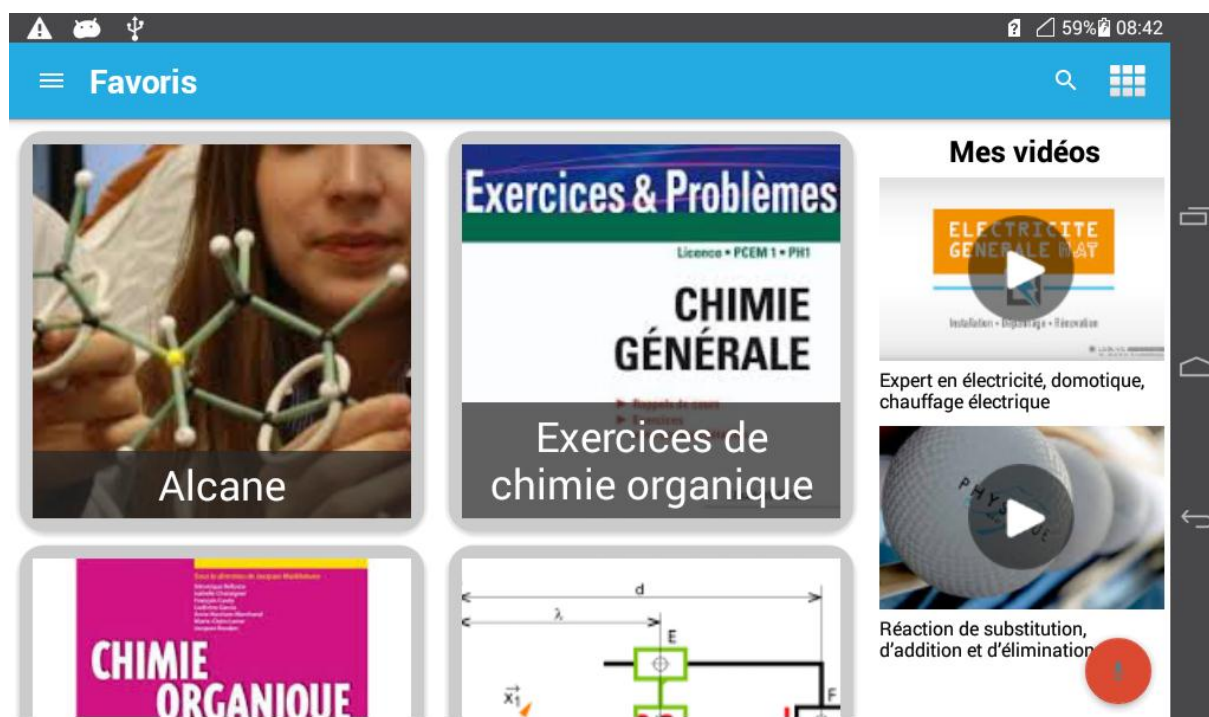


Figure 31 Ecran Favoris

Il affiche les items (Ici item symbolise une collection de livres comme le titre « **Alcane** » de la figure ci-dessus) favoris choisis par l'utilisateur. En cliquant sur le bouton « **microphone** », il peut commander via la voix la recherche de l'item désiré.

II.B.10. Ecran INFORMATION

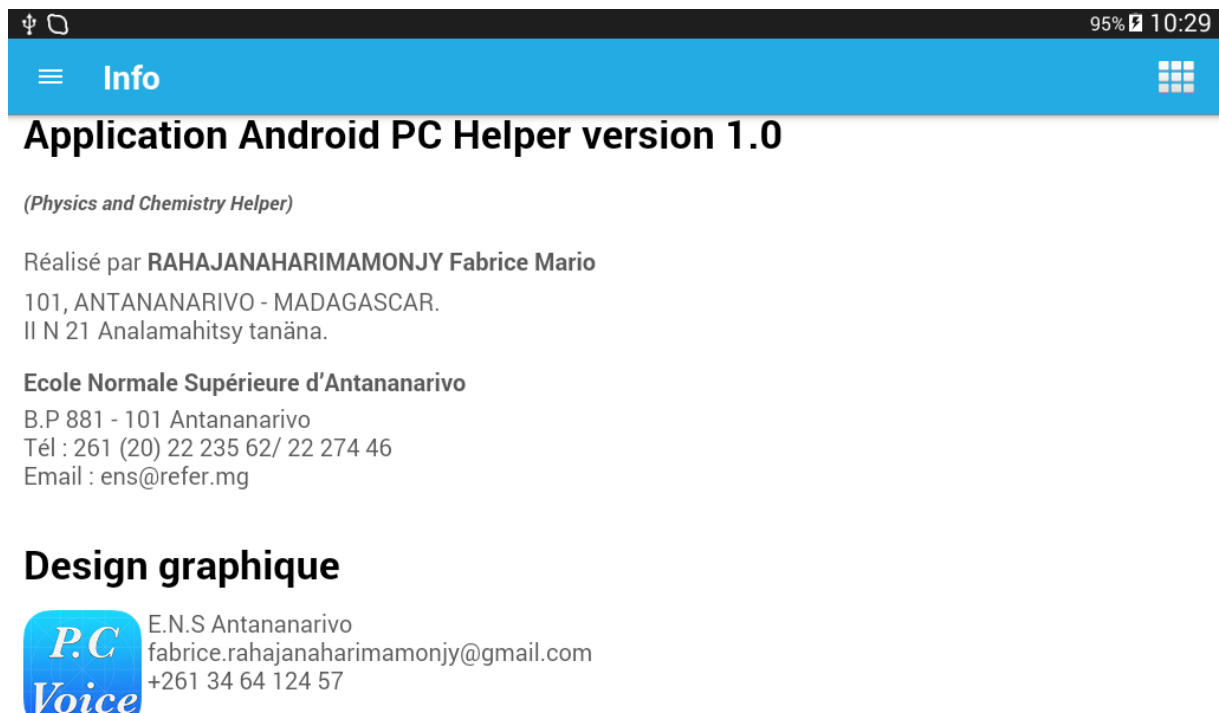


Figure 32 Ecran information

Il affiche les informations techniques de l'application, son développeur, son design graphique, ...

II.B.11. Ecran Checking Internet

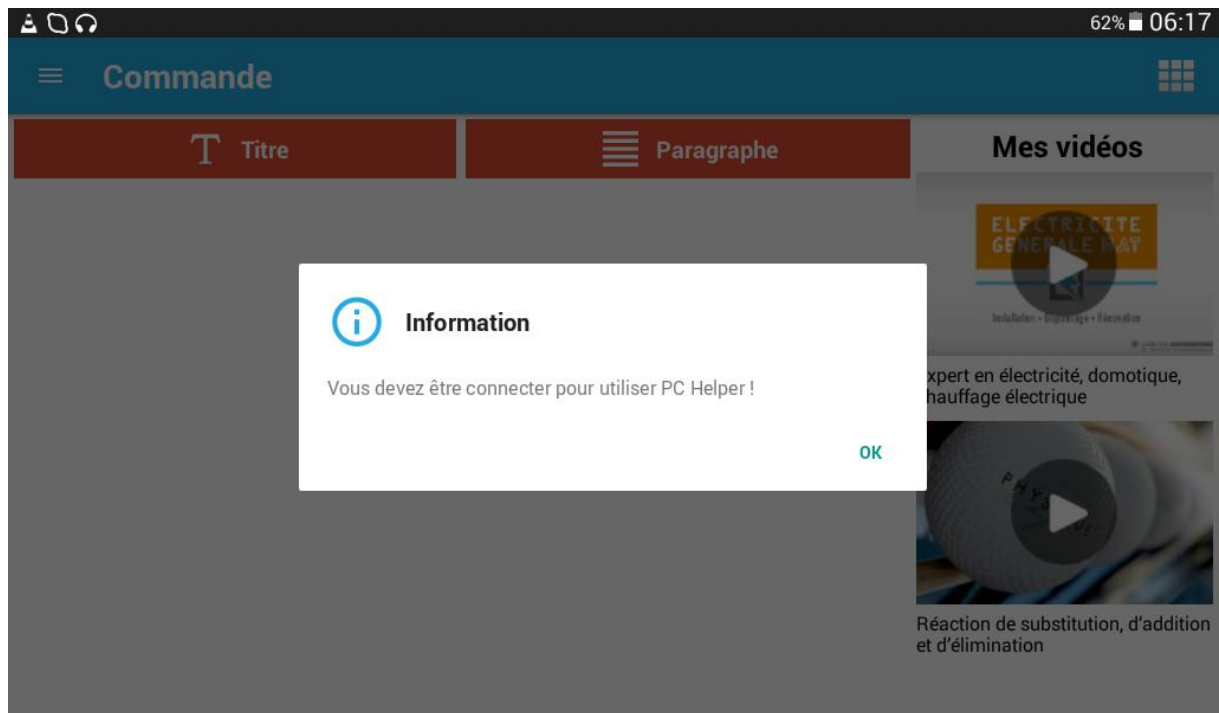


Figure 33 Ecran checking internet

Il détecte l'état de connexion internet du device (ici device est la Tablette). Le mode connecté (device connecté internet) permet d'enrichir le dictionnaire utilisé par le processus de la reconnaissance vocale c'est-à-dire pour avoir une meilleure performance de reconnaissance.

II.B.12. Ecran MICRO

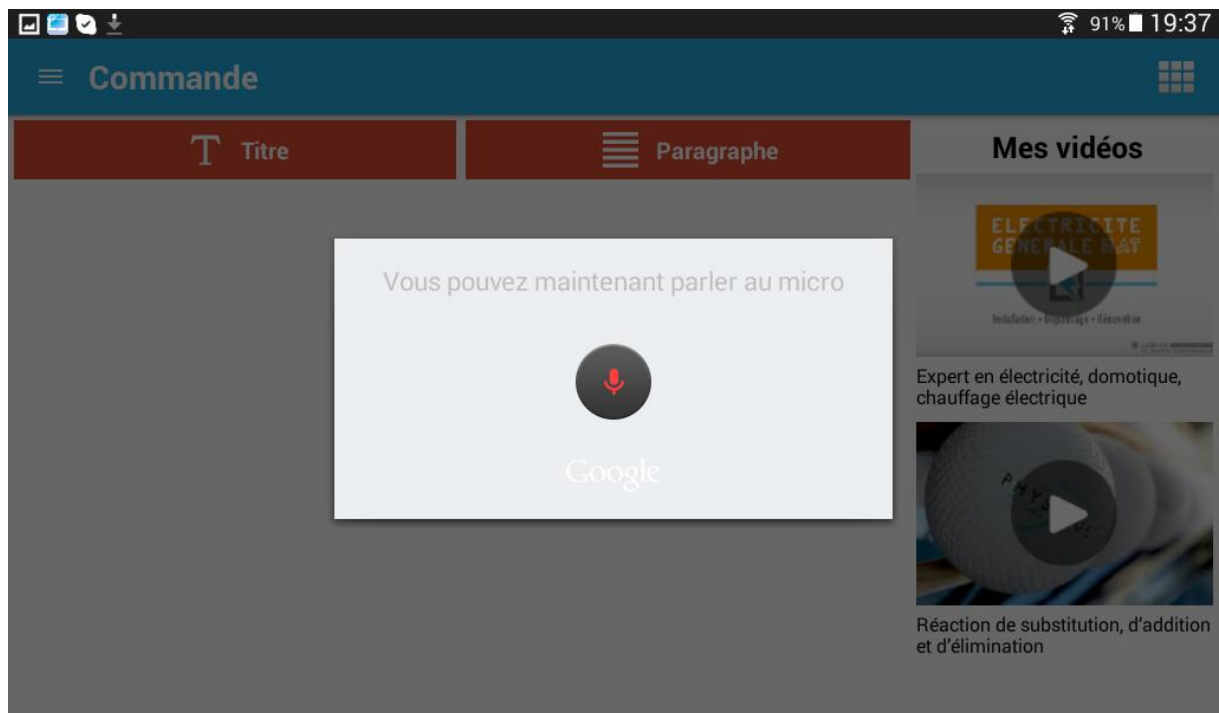


Figure 34 Ecran Micro

Il signale à l'utilisateur si l'application est prête à traiter la commande vocale. Il affiche un message afin que l'utilisateur sache à quel point il doit commencer sa commande. Les logiciels de reconnaissance vocale ont également besoin d'une période d'apprentissage. L'utilisateur doit utiliser micro-casque et lire un texte afin que le logiciel se familiarise avec sa voix.

Chapitre III. Les points forts de l'intégration de la reconnaissance vocale dans le secteur de l'éducation

Aujourd'hui, la grande majorité des écoles primaires et secondaires intègrent dans leur programme l'utilisation des tablettes numériques et le matériel informatique, à la fois comme technologie d'assistance pour les étudiants souffrant de handicaps physiques ou de difficultés d'apprentissage. L'intégration de la nouvelle technologie en effet, offre une opportunité pour les étudiants ayant un handicap d'accroître leur productivité.

La technologie nous a permis d'élaborer un logiciel de reconnaissance vocale. Pour la majorité des cas, la technologie nuise à l'enseignement, pour d'autre part celle-ci peut aussi porter de l'intérêt au niveau pédagogique.

En utilisant la reconnaissance vocale en classe par le biais des tablettes tactiles par exemple, l'enseignement/apprentissage, devient plus explicite mais aussi et surtout apporte des repères techniques pour les étudiants sourds et malentendants.

Certes, la reconnaissance vocale ne sait pas encore traiter les paroles spontanées, les hésitations, les reprises et les phrases dont la structure est éloignée de celle de la langue écrite, mais cela offre à ses utilisateurs un moyen d'exploiter le système et d'être plus vigilant vis-à-vis des résultats.

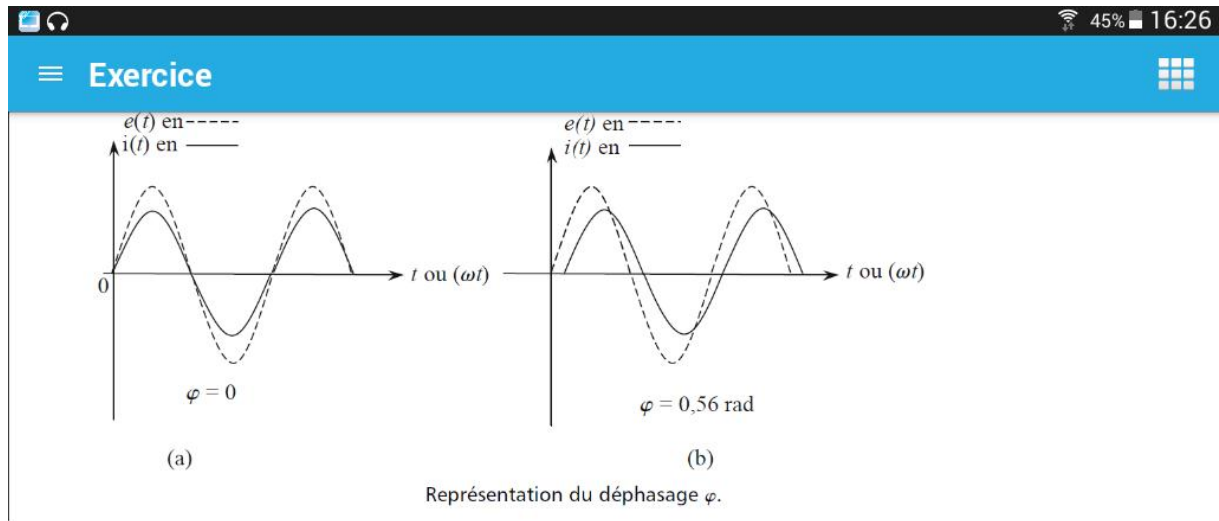
III.A. De réels avantages pour tous les étudiants

III.A.1. Rapidité

Grâce à l'utilisation de ce logiciel il est possible d'obtenir un résultat sans faire de calcul, il suffit d'entrer les données.

Prenons le cas de calcul de déphasage. Il illustre le fait quand on change la valeur de la résistance à $5\text{ k}\Omega$ et l'inductance 6 mH , on peut insérer via la commande vocale la valeur de la résistance ou de l'inductance voulue. L'application calcule automatiquement le déphasage correspondant à cette valeur dès que la commande est terminée. **(Figure 35)**

III.A.1.a. ECRAN D'EXERCICE D'APPLICATION



Resistance = 5.0 k

Inductance = 6.0 mH

Déphasage = 0,65 rad

Figure 35 Ecran d'exercice d'application

Les enseignants peuvent ajouter dans l'écran favori tous les cours avec leurs items (c'est-à-dire les titres de chaque cours), ainsi en cliquant sur le bouton microphone, il peut commander via la voix la recherche de l'item désiré. Cela permet une accessibilité rapide sur la leçon recherchée.

Les figures ci-dessous démontrent, une recherche du cours intitulé « oscillations ».

Avant la clique du bouton microphone

III.A.1.b. ECRAN D'APPLICATION FAVORIS



Figure 36 Ecrans d'application favoris 1

En cliquant sur le bouton microphone, et commander via la voix la recherche de l'item oscillation, On a :

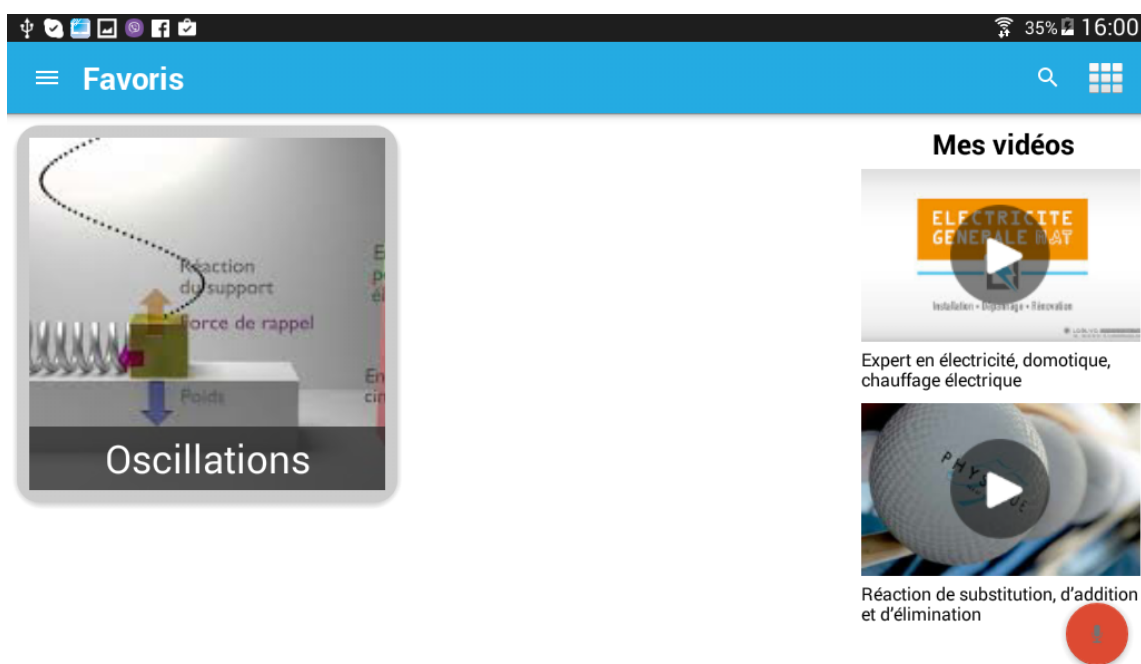


Figure 37 écrans d'application favoris 2

III.A.2. Prendre de notes :

L'application peut aussi être utilisée comme un carnet de notes. Aux travaux pratiques dans le laboratoire de chimie par exemple, lorsqu'un élève a les deux mains occupées, il peut utiliser sa voix pour prendre de note, au lieu d'écrire.

Prenons le cas d'un dosage acide base, l'élève peut en même temps faire la pratique et aussi notée sans prendre un stylo et cahier sur le volume de la base restante ou d'acides restants dans la burette de l'action grâce à l'enregistrement de l'application. (**Figure 38**)

III.A.2.a. ECRANS DE NOTES



Figure 38 Ecran de note 1

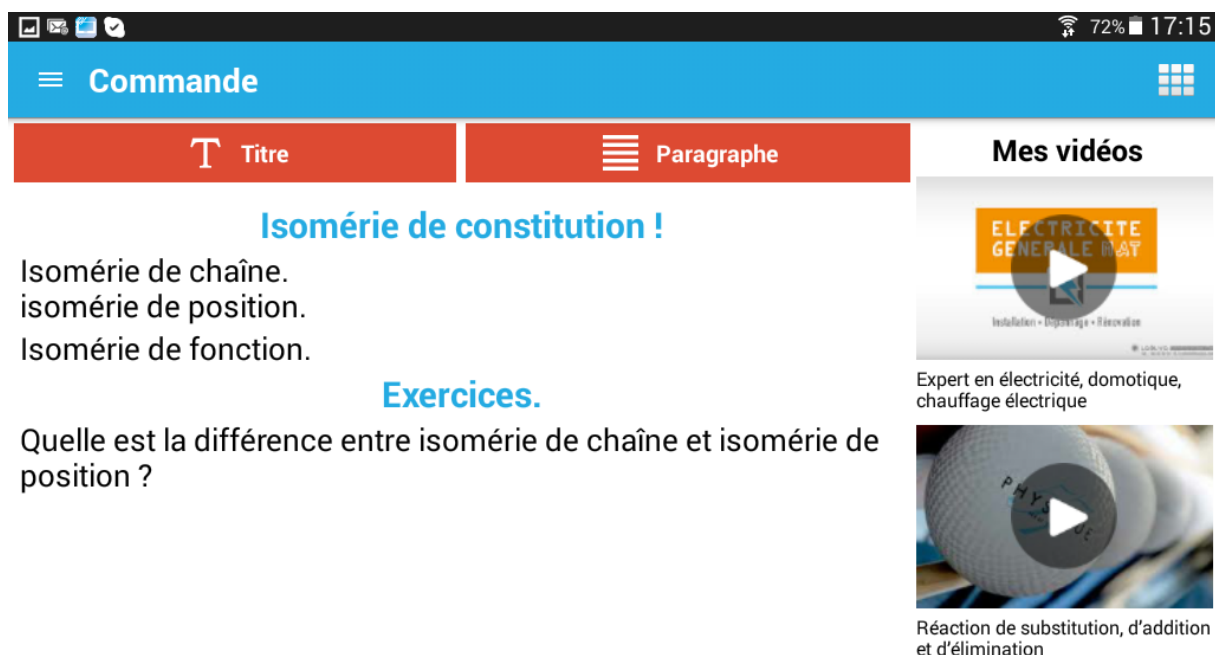


Figure 39 Ecran de note 2

L'utilisateur peut également préciser les signes de ponctuation « point », « virgule », « point d'exclamation », la « mise en page » et le « retour à la ligne ». Cas de la (**Figure 39**), lorsque vous dites « point d'interrogation », le logiciel insère la ponctuation correspondante.

Plus on exploite le logiciel, plus il devient « intelligent et cognitif ». En effet, il apprend et mémorise les caractéristiques de la voix.

Pour un meilleur résultat il est nécessaire de tenir compte de certains points :

- ✓ Il faut parler sans hésitation et bien articuler la prononciation.
- ✓ Il faut bien se préparer avant de parler et éviter d'autres bruits parasites

L'application « PC Helper » affichera correctement les mots prononcés sous forme de texte.

III.B. Accessibilité des étudiants souffrant d'un handicap physique

Les solutions de reconnaissance vocale comme « PC Helper » apportent une aide précieuse aux étudiants souffrant de toutes sortes de handicaps physiques comme la paralysie, la tétraplégie, les handicaps visuels, les étudiants sourds et malentendants et les douleurs liées à l'utilisation prolongée du clavier.

Globalement voici comment le « PC Helper » peut les aider :

- utilisation de la tablette en mains libres
- une aide pédagogique pour les étudiants sourds et malentendants

III.B.1. Utilisation de la tablette en mains libres

Les tablettes installées d'une application spécifique de reconnaissance vocale permettent de travailler les mains libres et de créer des documents, d'accéder à des données ou d'explorer un système de fichiers. Les solutions logicielles comme «PC Helper » offrent un moyen plus efficace de contrôler une tablette, moins pénible mais cognitif.

III.B.2. Une aide pédagogique pour les étudiants sourds et malentendants.

Actuellement, dans l'enseignement, la loi autorise les sourds et les jeunes malentendants à suivre les cours dans les mêmes salles de classe que les élèves entendant. C'est une opportunité pour ces élèves de poursuivre leurs études, n'empêche qu'ils rencontrent de grandes difficultés pour s'intégrer véritablement dans la classe. Le moyen de communication est une barrière qui les empêche d'apprendre correctement, demandera surtout l'utilisation de la langue des signes. En outre, le nombre d'enseignants capables de communiquer en langue des signes est insuffisant, ce qui rend la situation plus difficile.

Le développement de l'aide pédagogique avec « PC Helper » pour les étudiants sourds et malentendants est mis en œuvre afin d'apporter une réponse à ce problème.

Avec ce système, il sera possible pour ses élèves d'apprendre comme pour les élèves entendant. Ainsi, ils ne se sentent plus isolés et la notion communauté d'apprentissage prend alors un véritable sens. Chaque mot prononcé par l'enseignant est converti en texte par le logiciel de reconnaissance vocale et projeté en grands caractères sur l'écran dans la salle de classe. Cela permet aussi, aux sourds et malentendants de poser des questions aux enseignants pendant le cours.

« Lorsqu'ils utiliseront la reconnaissance vocale, on ne pourra plus faire la différence entre leur écriture et celle des autres et vice versa ».

III.C. Technologie d'assistance à l'usage des étudiants souffrant de difficultés d'apprentissage

Les ergothérapeutes accompagnent les personnes en situation de handicap pour qu'elles soient intégrées dans leur vie sociale, scolaire ou professionnelle en favorisant leur autonomie dans les gestes quotidiens. Les enfants porteurs de troubles spécifiques des apprentissages présentent des difficultés en classe il est important de les accompagner.

Les travaux de recherche et l'expérience pratique continuent à démontrer les potentialités qu'offre la reconnaissance vocale pour aider ses étudiants, que le cas soit dyslexie, dysgraphie, troubles de la mémoire et autres difficultés cognitives ces maladies qui nuisent à leur aptitude à la lecture, à l'expression écrite et à l'orthographe.

Les outils de reconnaissance vocale comme « PC Helper » peuvent développer l'envie d'écrire et d'apprendre chez des étudiants qui se trouvaient auparavant dans l'incapacité de produire des travaux écrits, en raison de troubles d'apprentissage. La technologie accomplit ce tour de force :

- en permettant aux étudiants de passer des idées au texte plus facilement.
- en contournant les aspects les plus frustrants de l'expression écrite, notamment la recherche d'idées, l'exposé et l'orthographe.

- en offrant une aide corrective en lecture tant pour le déchiffrage que la compréhension.

Grâce à « PC Helper », des étudiants ayant toutes sortes de difficultés d'apprentissage peuvent atteindre leur potentiel maximal. Le logiciel permet d'accomplir leur travail au son de la voix et améliore considérablement la production des étudiants atteints de troubles qui les empêchent de maîtriser la langue écrite.

III.C.1. Transformer les idées en textes écrits

L'utilisation d'un logiciel de reconnaissance vocale peut aider les étudiants présentant des difficultés d'apprentissage liées au langage à exprimer leurs idées sur le papier. En effet, ces étudiants éprouvent des difficultés à exprimer leurs pensées à l'écrit : leurs idées sont souvent moins développées et, ayant peur de commettre des erreurs, ils utilisent des vocabulaires trop simples. Certaines personnes ont tout simplement renoncé à écrire. Les outils de reconnaissance vocale peuvent être d'une aide précieuse pour ces personnes afin qu'ils voient les mots s'afficher à l'écran à mesure qu'ils les dictent. Cette approche « dire-et-voir » aide les étudiants à mieux saisir la relation entre la graphie d'un mot et sa prononciation. Autre avantage non négligeable, de l'approche une technique permettant aux enfants confrontés à l'échec à plusieurs reprises de reprendre confiance en eux.

III.C.2. Permettre une plus grande autonomie

Par le passé, de nombreux étudiants ayant des difficultés d'apprentissage dépendaient d'un tiers comme ses parents, ses enseignants, ses formateurs ou ses amis pour transcrire leurs rédactions et leurs devoirs.

Cette approche traditionnelle présente toutefois de sérieux inconvénients. Tout d'abord, elle rend l'étudiant dépendant d'un tiers pour effectuer ses travaux scolaires. En outre, elle nuit au développement des compétences en expression écrite ; sans brouillon à lire

ou relire au fur et à mesure du développement de chaque phrase, il manque à l'étudiant le sens global du flux d'écriture.

L'application de reconnaissance vocale offre une alternative rationnelle à la tierce personne. Il permet aux étudiants de parler devant leur tablette à un rythme naturel et de regarder les mots prononcés s'affichant à l'écran.

Ces étudiants sont désormais en mesure d'utiliser la reconnaissance vocale pour atteindre de nouveaux niveaux d'autonomie en écriture, lecture et apprentissage.

CONCLUSION

Au cours de ce travail, nous avons pu relever les performances du système de reconnaissance vocale sur le traitement de la parole et sur la méthode de comparaison dynamique ou (DTW).

Le système de reconnaissance vocale est devenu plus pratique depuis quelque temps. Il est utilisé dans tous les domaines professionnels et enseignements. Le système de la reconnaissance est surtout basé sur la voix. Elle est une technique informatique permettant l'analyse automatique de la parole. Un microphone permet à la tablette tactile, à l'ordinateur, ou au smartphone de capter la parole, de l'analyser et de la transcrire sous la forme d'un texte exploitable.

Sur le plan de recherche, l'intégration de reconnaissance vocale apporte des points forts dans le secteur de l'éducation. La reconnaissance vocale constitue un grand atout pour les élèves surtout ceux qui ont certains handicaps, pour notre cas ceux qui sont sourds et malentendants.

Notre objectif est d'apporter des solutions qui facilitent et améliorent le processus d'apprentissage, par le logiciel reconnaissance vocale. Il permet aux jeunes handicapés d'étudier et de communiquer avec les élevés normaux ainsi que les enseignants, tout en étant dans la même classe et ayant la même vitesse d'apprentissage.

L'enseignement des élèves présentant des handicaps avec ceux qui ne le sont pas demande un travail supplémentaire pour les enseignants du fait qu'ils doivent apprendre la langue des signes. Mais grâce au logiciel, les enseignants sont surtout menés à se concentrer sur la conception de cours et au meilleur moyen d'apprentissage à toute personne.

Les tablettes tactiles, depuis quelque temps, sont intégrées comme un outil pédagogique et présent dans tous les lycées. L'exploitation du logiciel de reconnaissance vocale est alors accessible à tout membre de l'enseignement, et peut être utilisée au niveau de l'administration, de l'enseignement, et également des élèves.

Le logiciel élaboré est loin d'être parfait, il y aura toujours des versions améliorées. Nous encourageons vivement nos cadets dans la maîtrise et l'exploitation des logiciels récents, de l'améliorer et même de le rénover afin d'être présenté dans le cadre de la préparation des mémoires de fin d'études à l'ENS.

Bibliographies :

- [1]. Bonnet, F., & Devèse, B. (2004). *La reconnaissance automatique de la parole, Calculatrice vocale*. EPITA.
- [2]. Calliope. (1989). *La parole et son traitement automatique* (éd. Masson).
- [3]. DALLA CORTE, L. (2012). *Reconnaissance vocale sur Smartphone par apprentissage automatique*. 2ème Master en Sciences Informatiques, Faculté des Sciences Appliquées de l'Université de Liège.
- [4]. Didier , M. (2000). *L'apport d'une approche automatique*. Thèse de doctorat, institut de police scientifique et de criminologie de l'université de Lausanne.
- [5]. Jean-Paul, A. (1987). *Théorie et technique des images et des sons*. Tome 1.
- [6]. Rodolphe , B. (1998). *Reconnaissance vocale*. Examen Probatoire pour l'obtention du Diplôme d'Ingénieur,C.N.A.M.
- [7]. Saporta, G. (s.d.). *Probabilités, analyse des données et statistique* (éd. 1990).
- [8]. Scherrer, B., & Gaëtan, M. (1984). *Biostatistique* (éd. C.P.965). Chicoutimi, Québec, Canada.
- [9]. Solan, L., & Tiersma, P. (2003). *Falling on Deaf Ears*. Legal Affairs NOV-DEC.
- [10]. Yarmey, A. (1995). *Earwitness Speaker Identification* (Vol. I). Psychology,Public Policy,and Law.
- [11]. DALLA CORTE, L. (2012). *Reconnaissance vocale sur Smartphone par apprentissage automatique* .2ème Master en Sciences Informatiques, Faculté des Sciences Appliquées de l'Université de Liège.
- [12]. E341. *Théorie du Signal*. Cours en 3ème année (2009), Ecole Supérieur Polytechnique d'Antananarivo-Département Electronique.
- [13]. RAZAFIARISON Zo Manankasina. (2009). *Réalisation d'une suite complète d'acquisition, de génération et de traitement du son*. Livre de mémoire

Webographie :

- [14]. <http://tcts.fpms.ac.be/couvreur> « ON THE USE OF PHASE INFORMATION FOR SPEECH RECOGNITION» récupérée le 08 mai 2016
- [15]. www-clips.imag.fr/geod/User/laurent.besacier/M2R-ILP/1.a. « Parole » du 13 mai 2016
- [16]. http://www.irit.fr/ACTIVITES/EQ_ARTPS « recherche des composantes parole » récupérée le 08 juillet 2016
- [17]. <http://r.battault.free.fr/probatoire/probatoire.html> « La Reconnaissance vocale » récupérée le 10 juillet 2016
- [18]. <http://www.vieartificielle.com/index.php?action=article&id=191> « traitement du signal pour la reconnaissance vocale » récupérée le 10 juillet 2016
- [19]. <http://www.chaumetsoftware.com/timbre.html> « Le timbre de la voix » récupérée le 20 juillet 2016
- [20]. <http://www.evariste.org/100tc/1996/f057.html> - « Reconnaissance de la parole » récupérée le 08 mai 2016
- [21]. <http://www.cavi.univ-paris3.fr/ilpga/ed/dr/jvdr2/articlesJV> « la musicalité de la voix parlée » récupérée le 08 juillet 2016
- [22]. <http://ww2.acpoitiers.fr/ecoles/IMG/pdf/> « outils facilitateurs scolarité » récupérée le 17 juillet 2016
- [23]. <http://campus.lemonde.fr/bac-s/fiche/> « emetteurs-et-recepteurs-sonores_2753 » récupérée le 19 Aout 2016

Titre : « Technique d'enseignement via une application de reconnaissance vocale »

Auteur : RAHANJANAHARIMAMONJY Fabrice Mario

RESUME :

Le traitement de la parole permet d'établir les différents paramètres du signal vocal. Les méthodes utilisées dans ce travail ont été l'analyse spectrale de la transformée de Fourier discrète, le spectrogramme, le codage, l'échelle de Mels. Les paramètres obtenus à partir de ce traitement de parole sont nécessaires pour l'élaboration de la reconnaissance vocale. La technique de comparaison utilisée pour la reconnaissance vocale est la déformation temporelle dynamique ou DTW.

Le deuxième chapitre est centré sur la présentation d'un logiciel didacticiel « PC Helper » qui vise à faciliter l'enseignement /apprentissage. En voyant dans ce chapitre la programmation et les interfaces de « PC-Helper »

Le dernier chapitre résume les points forts d'intégration de la reconnaissance vocale dans le secteur de l'éducation, ce dernier aborde des réels avantages pour tous types d'étudiants. L'application est accessible à tous, qu'ils souffrent d'un handicap ou non. Aussi, la nouvelle application offre une technologie d'assistance à l'usage des étudiants ayant des difficultés ou complexes d'apprentissage.

Nombre de pages : 68

Nombre de figures : 39

Nombre de tableaux : 04

Mots clés : Reconnaissance vocale, Numérisation, TF, Android

Adresse de l'auteur :

Lot : II N 21 A Analamahitsy-Tana 101-Madagascar

Tél : +261 34 64 124 57

e-mail : fabrice.rahajanaharimamonjy@gmail.com