Résumé

L’assistance aux apprenants du vocabulaire arabe et la correction automatique des pathologies vocales présentent des tâches pertinentes pour plusieurs applications de traitement automatique de la parole telles que l’évaluation des prononciations, la reconnaissance automatique de la parole, la dictée vocale, etc. C’est dans ce contexte que s’inscrit ce projet de fin d’études qui propose une application mobile pour la détection des pathologies vocales contenues dans le discours arabe suivant une approche phonétique-probabiliste. Notre travail présente une initiative, qui rend la technologie accessible pour tous.

Abstract

Arabic vocabulary assistants and automatic speech correction provide relevant tasks for several automatic speech processing applications such as pronunciation evaluation, automatic speech recognition, voice dictation, etc. It is in this context that this end-of-study project is proposed which proposes a mobile application for the detection of vocal pathologies contained in the Arabic discourse following a phonetic-probabilistic approach. Our work presents an initiative that makes technology accessible to all.

**نبذة مختصرة**

يوفر مساعدي المفردات العربية وتصحيح الكلام التلقائي، مهام ذات صلة للعديد من تطبيقات معالجة الكلام التلقائية، مثل تقييم النطق ، والتعرف التلقائي على الكلام ، والإملاء الصوتي ، إلخ. في هذا السياق ، يُقترح هذا المشروع تطبيقًا للهاتف الذكي لاكتشاف المشاكل الصوتية الواردة في الخطاب العربي. يقدم عملنا مبادرة تجعل التكنولوجيا في متناول الجميع.

Remerciements

Je tiens à remercier tout d’abord mon encadrant M. Naim TERBEH, je serais vaniteux si je devais énumérer en ces quelques lignes votre remarquables qualités humaines et professionnelles, veuillez trouver ici l’expression et le témoignage de mon gratitude ressentie. Je tiens d’emblée à vous remercier spécialement de tout l’intérêt que vous aviez bien voulu porter à mon travail, vos conseils, vos explications, vos directions et remarques, que ce modeste travail vous honore et témoigne ma reconnaissance.

Je tiens aussi à remercier tous les membres du jury pour avoir accepté l’évaluation de mon travail.

J’exprime toute ma reconnaissance et gratitude à l’administration et à l’ensemble du corps professionnels de la faculté des sciences de Monastir, pour leurs efforts qu’ils ont fournis afin de garantir la continuité et l’aboutissement de ce projet de fin d’études.

Enfin, je remercie tous ceux qui, de près ou de loin, ont contribué à la réalisation de ce modeste travail, en particulier ma famille et mes amis qui par leurs prières et leurs encouragements, mon permis à surmonter tous les obstacles.

Table des matières

[Introduction Générale 1](#_Toc12467667)

[Chapitre 1 : La reconnaissance automatique de la parole 3](#_Toc12467668)

[1. Introduction 4](#_Toc12467669)

[2. Définition 4](#_Toc12467670)

[3. Historique 4](#_Toc12467671)

[3.1 La naissance 4](#_Toc12467672)

[3.2 Les premiers mots 4](#_Toc12467673)

[3.3 La reconnaissance du langage 5](#_Toc12467674)

[4. Principe de fonctionnement 5](#_Toc12467675)

[4.1 Problématique 5](#_Toc12467676)

[4.2 Fonctionnement 6](#_Toc12467677)

[4.2.1 Reconnaissance par comparaison à des exemples 7](#_Toc12467678)

[4.2.2 Reconnaissance par modélisation d’unités de parole 7](#_Toc12467679)

[5. Application 8](#_Toc12467680)

[5.1 Services vocaux 8](#_Toc12467681)

[5.2 Contrôle de qualité, saisie des données 9](#_Toc12467682)

[5.3 Avionique 9](#_Toc12467683)

[5.4 Formation 9](#_Toc12467684)

[5.6 Aide aux handicapés 9](#_Toc12467685)

[5.7 Autres applications 10](#_Toc12467686)

[6. Problématiques 10](#_Toc12467687)

[7. Difficultés d’ordre pathologique 10](#_Toc12467688)

[7.1 Parole saine 11](#_Toc12467689)

[7.2 Parole pathologique 11](#_Toc12467690)

[7.3 Typologie des pathologies vocales 12](#_Toc12467691)

[7.4 Facteurs sources des pathologies vocales 13](#_Toc12467692)

[8. Conclusion 13](#_Toc12467693)

[Chapitre 2 : Choix Méthodologique et Environnement Technique 15](#_Toc12467694)

[1. Introduction 16](#_Toc12467695)

[I. La méthode de détection des pathologies vocales 16](#_Toc12467696)

[1. Introduction 16](#_Toc12467697)

[2. Génération d’un modèle phonétique 16](#_Toc12467698)

[2.1 Préparation des bases de langage arabe 16](#_Toc12467699)

[2.2 Nettoyage des corpus 17](#_Toc12467700)

[2.3 Modélisation phonétique 17](#_Toc12467701)

[3. Approche proposée 17](#_Toc12467702)

[3.1 Détection des pathologies vocales 17](#_Toc12467703)

[3.1.1 Produit scalaire dans 17](#_Toc12467704)

[3.1.2 Distance phonétique et classification de la parole 18](#_Toc12467705)

[4. Récapitulation 19](#_Toc12467706)

[II. Environnement technique 19](#_Toc12467707)

[1. Python 19](#_Toc12467708)

[2. NumPy 19](#_Toc12467709)

[3. Flask 20](#_Toc12467710)

[4. Flask-RESTful 20](#_Toc12467711)

[5. Java 20](#_Toc12467712)

[6. Android 21](#_Toc12467713)

[7. PocketSphinx 21](#_Toc12467714)

[8. PyCharm 21](#_Toc12467715)

[9. Android Studio 21](#_Toc12467716)

[10. Postman 22](#_Toc12467717)

[11. Visual Studio 23](#_Toc12467718)

[III. Conclusion 23](#_Toc12467719)

[Chapitre 3 : Réalisation 24](#_Toc12467720)

[1. Introduction 25](#_Toc12467721)

[2. Le service de détection des pathologies vocales dans la parole Arabe 25](#_Toc12467722)

[2.1 Les services web 25](#_Toc12467723)

[2.1.1 Types des services web 25](#_Toc12467724)

[3. Déploiement du service 26](#_Toc12467725)

[4. Le modèle acoustique 40](#_Toc12467726)

[4.1 Pourquoi entrainer un modèle acoustique ? 40](#_Toc12467727)

[4.2 Entrainement du modèle acoustique 41](#_Toc12467728)

[5. L’application mobile 42](#_Toc12467729)

[Glossaire 48](#_Toc12467730)

[Liste bibliographique 49](#_Toc12467731)

Liste des figures

[Figure 1 - Site web de Python 19](#_Toc12466889)

[Figure 2 - Site Web Flask "http://flask.pocoo.org/" 20](#_Toc12466890)

[Figure 3 - Site Web FalskRESTful https://flask-restful.readthedocs.io/en/latest/ 20](#_Toc12466891)

[Figure 4 - PyCharm 21](#_Toc12466892)

[Figure 5 - Android Studio 22](#_Toc12466893)

[Figure 6 - Postman 22](#_Toc12466894)

[Figure 7 - Visual Studio 23](#_Toc12466895)

[Figure 8 - Description abstraite de la communication entre l’application mobile et l’API REST 25](#_Toc12466896)

[Figure 9 - Premiers accès au serveur 26](#_Toc12466897)

[Figure 10 - Vérification des mis à jour 27](#_Toc12466898)

[Figure 11 - Ajout d'un nouvel utilisateur 27](file:///D:\New%20folder%20(3)\rapport-draft%20-%202.docx#_Toc12466899)

[Figure 12 - Donner au nouvel utilisateur les privilèges root 28](#_Toc12466900)

[Figure 13 - Modification du service ssh 29](#_Toc12466901)

[Figure 14 - Redémarrage du service ssh 29](#_Toc12466902)

[Figure 15 - Reconnecter au server en tant qu’utilisateur hamza 30](#_Toc12466903)

[Figure 16 - Installation de nginx 30](#_Toc12466904)

[Figure 17 - État du pare-feu 31](#_Toc12466905)

[Figure 18 - Autorisation du ssh à travers le pare-feu 32](#_Toc12466906)

[Figure 19 - État du service nginx 32](#_Toc12466907)

[Figure 20 - Fichier de configuration de nginx 33](#_Toc12466908)

[Figure 21 - Création d'un lien symbolique vers le fichier de configuration 33](#_Toc12466909)

[Figure 22 - Clonage du code 34](#_Toc12466910)

[Figure 23 - Installation du package build-essential 35](#_Toc12466911)

[Figure 24 - Installation du package python3.7 et python3.7-dev 35](#_Toc12466912)

[Figure 25 - Création d'un environnement virtuel 36](#_Toc12466913)

[Figure 26 - Activation de l'environnement virtuel 36](#_Toc12466914)

[Figure 27 - Installation de Flask, FlaskRESTful, uwsgi, numpy 37](#_Toc12466915)

[Figure 28 - Fichier de configuration de service 37](#_Toc12466916)

[Figure 29 - Fichier de configuration de uwsgi 38](#_Toc12466917)

[Figure 30 - Démarrage et vérification de l'état du service 38](#_Toc12466918)

[Figure 31 - Première test du service de détection des pathologies 39](#_Toc12466919)

[Figure 32 - Deuxième test du service de détection des pathologies 39](#_Toc12466920)

[Figure 33 - Troisième test du service de détection des pathologies 40](#_Toc12466921)

[Figure 34 - La liste des modèles acoustiques disponibles 41](#_Toc12466922)

[Figure 35 - Première exécution de l'application 42](file:///D:\New%20folder%20(3)\rapport-draft%20-%202.docx#_Toc12466923)

[Figure 36 - Application prête à être utiliser 42](file:///D:\New%20folder%20(3)\rapport-draft%20-%202.docx#_Toc12466924)

[Figure 37 - Application en écoute 43](file:///D:\New%20folder%20(3)\rapport-draft%20-%202.docx#_Toc12466925)

[Figure 38 - Aucun parole détecté 43](file:///D:\New%20folder%20(3)\rapport-draft%20-%202.docx#_Toc12466926)

[Figure 40 – Résultat obtenu (cas de la parole pathologique) 44](file:///D:\New%20folder%20(3)\rapport-draft%20-%202.docx#_Toc12466927)

[Figure 39 -Résultat obtenu (cas de la parole saine) 44](file:///D:\New%20folder%20(3)\rapport-draft%20-%202.docx#_Toc12466928)

# Introduction Générale

La production de la parole est un acte moteur complexe qui implique un grand nombre de muscles, de variables physiologiques et un contrôle neurologique impliquant différentes régions corticales et sous corticales. Nous distinguons trois systèmes intervenant à la production de la parole : le système respiratoire, le système laryngé et le système supra-laryngé (les articulateurs).

L’analyse des troubles de la parole demeure essentiellement clinique. Les mesures instrumentales sont peu répandues en pratique clinique. Les plus utilisées sont les mesures acoustiques et aérodynamiques. L’analyse automatique de la parole est une tâche complexe et a été longtemps négligée. Il en résulte une difficulté pour analyser dans la littérature l’effet des différents traitements (médicaux ou chirurgicaux). En effet, beaucoup d’études ne rapportent pas d’analyse spécifique de la parole. D’autre part, on assiste souvent à une confusion entre les modifications de la motricité orofaciale et la qualité de la parole qui reste l’objectif clinique.

L'évaluation des caractéristiques d'un désordre de la voix et l'effet que le désordre porte sur la capacité d'un locuteur de communiquer est un pas crucial pour concevoir un programme de sa gestion. Un processus de l'évaluation prospère permet orthophonistes de diagnostiquer le désordre de la voix, de déterminer l'efficacité relative de plusieurs méthodes de traitement et de formuler un pronostic.

Les techniques souvent utilisées par les médecins, pour analyser les symptômes des pathologies vocales, sont envahissantes comme l’endoscopie. Cependant, il est possible d'identifier les pathologies vocales en analysant certains paramètres du signal de la parole.

Nous, nous proposons dans ce projet une application mobile pour la détection des pathologies vocales contenues dans la parole arabe, suivant une modélisation phonétique-probabiliste. L’application développée est limitée à la classification de la parole produite en saine ou pathologique, et non pas à la localisation de la pathologie, ni à la typologie des pathologies vocales.

Dans ce travail on repondéra à la question : Est-ce que la parole produite est saine ou pathologique ?

# Chapitre 1 : La reconnaissance automatique de la parole

## 1. Introduction

La reconnaissance automatique de la parole est un domaine de la science ayant toujours eu un grand attrait auprès des chercheurs comme auprès du grand public. L’homme étant par nature paresseux, une telle technologie a toujours suscité chez lui une part d’envie et d’intérêt, ce que peu d’autres technologies ont réussi à faire.

Le domaine de la reconnaissance automatique de la parole est en pleine croissance et nous verrons dans ce chapitre que la technologie actuelle est très aboutie, pouvant commencer à répondre aux attentes de l’homme. Bien que des progrès soient encore à faire sur les systèmes complexes de reconnaissance, il est à noter que la reconnaissance de petits vocabulaires est quasiment parfaite, ce qui suffit largement pour des outils de traitements vocaux du quotidien.

## 2. Définition

La reconnaissance automatique de la parole est l’un des deux domaines du traitement automatique de la parole, l’autre étant la synthèse vocale. La reconnaissance automatique de la parole permet à la machine de comprendre et de traiter des informations fournies oralement par un locuteur. Elle consiste à employer des techniques d'appariement afin de comparer une onde sonore à un ensemble d'échantillons, composé généralement de mots mais aussi, plus récemment, de phonèmes (unité sonore minimale).

Ces deux axes et notamment la reconnaissance vocale, font appel aux connaissances de plusieurs sciences : l'anatomie (les fonctions de l'appareil phonatoire et de l'oreille), les signaux émis par la parole, la phonétique, le traitement du signal, la linguistique, l'informatique, l'intelligence artificielle et les statistiques.

Le traitement automatique de la parole ouvre des perspectives nouvelles compte tenu de la différence considérable existant entre la commande manuelle et vocale. L'utilisation du langage naturel dans le dialogue homme/machine met la technologie à la portée de tous et entraîne sa vulgarisation, en réduisant les contraintes de l'usage des claviers, souris et codes de commandes à maîtriser. En simplifiant le protocole de dialogue homme/machine, le traitement automatique de la parole vise donc aussi un gain de productivité puisque c'est la machine qui s'adapte à l'homme pour communiquer, et non l'inverse.

## 3. Historique

### 3.1 La naissance

Les premières tentatives de création d’une machine capable de comprendre le discours humain eurent lieu aux USA à la fin des années 40, au sein du Ministère de la Défense américain. Le but était de traduire et d’interpréter des messages russes interceptés.

### 3.2 Les premiers mots

Vers les années 1950 apparaît le premier système de reconnaissance de chiffres ; appareil entièrement câblé et très imparfait. En 1951, S.P. Smith présente un détecteur de phonèmes. Une année après, K.H. Davis, R Biddulph et S.Baleshek annoncent la première machine à aborder la reconnaissance de manière globale : les dix chiffres «zéro» à «neuf» sont reconnus analogiquement avec un bon taux de réussite pour une seule voix.

Puis en 1960, P.B. Denes et M.V. Matthews, pour reconnaître les dix premiers chiffres, comparent globalement les représentations temps fréquence, numérisées et normalisées en durée totale : le taux d'erreur est nul pour un seul locuteur et s’élève à 6% pour cinq locuteurs ayant participé à un apprentissage. H.F. Olson et H.Belar envisagent, en 1961, la reconnaissance d'unités phonétiques autres que les phonèmes : leurs unités sont des «syllabes phonétiques» que le locuteur doit articuler séparément ou, du moins, avec une chute importante du niveau sonore en guise de séparation ; il s'agit donc presque d'une reconnaissance par mots, étant entendu que ces «mots» sont courts et que leur répertoire est limité : 2000 syllabes suffisent à couvrir 98% des besoins de la langue anglaise. J.Dreyfus-Graf met au point en 1961 son «phonétographe», appareillage analogique composé de vingt filtres passe-bande et de circuits identificateurs de phonèmes. Le phonétographe utilise des « compresseurs sélectifs » qui augmentent l'émergence de certains sons. Obtenu en temps réel, le résultat est spectaculaire ; cependant l'appareil ne fonctionne qu'avec un seul locuteur qui doit adapter sa diction à la machine : hauteur, intensité, rythmes très faibles.

Après avoir constaté que l'identification des phonèmes dans le signal de parole est un problème beaucoup plus compliqué, les chercheurs se tournent, entre 1965 et 1970, d'une part vers la reconnaissance par mots isolés en vue d'applications pratiques comme la commande vocale, d'autre part vers l'utilisation d'informations de niveau linguistique supérieur avec lexique et syntaxe, pour compléter le message vocal reconnu au niveau phonétique. Cette seconde approche prend le nom, quelque peu abusif, de « compréhension automatique de la parole ».

### 3.3 La reconnaissance du langage

Dès lors, les recherches dans le domaine de la reconnaissance vocale n’ont cessé de progresser dans le sens de la compréhension du langage parlé et des phrases structurées. Aujourd’hui, le taux d’erreur ainsi que le temps d’apprentissage des systèmes de reconnaissance ne cessent de diminuer pour atteindre de nos jours des résultats proche de 95%. Ce taux est évidement variable selon la difficulté du langage. En effet la machine a parfois du mal à éviter certains pièges linguistiques.

## 4. Principe de fonctionnement

Dans cette partie on va expliquer comment fonctionnent les principaux systèmes de reconnaissance vocale. Néanmoins, le sujet de ce rapport n’étant pas une étude approfondie des techniques de reconnaissance vocale, les principes mathématiques et de traitement du signal ne seront pas traités mais juste évoqués dans ce chapitre.

### 4.1 Problématique

Pour bien appréhender le problème de la reconnaissance automatique de la parole, il est bon d'en comprendre les différents niveaux de complexité et les différents facteurs qui en font un problème difficile.

* **Le système doit-il être optimisé pour un unique locuteur ou est-il destiné à devoir se confronter à plusieurs utilisateurs ?**

On peut aisément comprendre que les systèmes dépendants d’un seul locuteur sont plus faciles à développer et sont caractérisés par de meilleurs taux de reconnaissance que les systèmes indépendants du locuteur étant donné que la variabilité du signal de parole est plus limitée. Cette dépendance au locuteur est cependant acquise au prix d'un entraînement spécifique à chaque utilisateur. Ceci n'est néanmoins pas toujours possible. Par exemple, dans le cas d'applications téléphoniques, on comprend bien que les systèmes puissent être utilisés par n'importe qui et donc être indépendants du locuteur.

Bien que la méthodologie de base reste la même, cette indépendance au locuteur est obtenue par l'acquisition de nombreux locuteurs (couvrant si possible les différents dialectes) qui sont utilisés simultanément pour l'entraînement de modèles susceptibles d'en extraire toutes les caractéristiques majeures. Une solution intermédiaire parfois utilisée, consiste à développer des systèmes capables de s'adapter rapidement (de façon supervisée ou non) aux nouveaux locuteurs.

* **Le système reconnaît-il des mots isolés ou de la parole en continue ?**

Evidemment, il est plus simple de reconnaître des mots isolés bien séparés par des périodes de silence que la séquence de mots constituant une phrase. En effet, dans ce dernier cas, non seulement la frontière entre les mots n'est plus connue mais les mots deviennent fortement articulés.

Dans le cas de la parole continue, le niveau de complexité varie également selon qu'il s'agisse de texte lu, de texte parlé ou, beaucoup plus difficile, de langage naturel avec ses hésitations, phrases grammaticalement incorrectes, faux départs, etc… Un autre problème, qui commence à être bien maîtrisé, concerne la reconnaissance de mots clés en parole libre. Dans ce dernier cas, le vocabulaire à reconnaître est relativement petit et bien défini mais le locuteur n'est pas contraint de parler en mots isolés. Par exemple, si un utilisateur est invité à répondre par « oui » ou « non », il peut répondre « oui, s'il vous plaît ». Dans ce contexte, un problème qui reste particulièrement difficile est le rejet de phrases ne contenant aucun mot clé.

* **Le système est-il robuste ?**

Autrement dit, le système est-il capable de fonctionner proprement dans des conditions difficiles ? En effet, de nombreuses variables pouvant affecter significativement les performances des systèmes de reconnaissance ont été identifiées :

* + Bruits d'environnement (dans une rue, un bistrot etc…)
  + Déformation de la voix par l’environnement (réverbérations, échos, etc…)
  + Qualité du matériel utilisé (micro, carte son etc…)
  + Bande passante fréquentielle limitée (fréquence limitée d’une ligne téléphonique)
  + Elocution inhabituelle ou altérée (stress, émotions, fatigue, etc…)

Certains systèmes peuvent être plus robustes que d'autres à l'une ou l'autre de ces perturbations, mais en général, les systèmes de reconnaissance de la parole sont encore sensibles à ces perturbations.

### 4.2 Fonctionnement

Le problème de la reconnaissance automatique de la parole consiste à extraire l'information contenue dans un signal de parole, typiquement par échantillonnage du signal électrique obtenu à la sortie d’un microphone, afin qu’il puisse être comparé à des modèles sous forme numérique. Parmi plusieurs techniques de reconnaissance, il y en a deux qui sont majoritairement utilisées afin de parvenir à résoudre ce problème : la comparaison à des exemples et la comparaison d’unités de parole.

#### 4.2.1 Reconnaissance par comparaison à des exemples

L’idée, très simple dans son principe, consiste à faire prononcer un ou plusieurs exemples de chacun des mots susceptibles d’être reconnus, et à les enregistrer sous forme de vecteurs acoustiques (représentation numérique du signal sonore).

Puisque cette suite de vecteurs acoustiques caractérise complètement l’évolution de l’enveloppe spectrale du signal enregistré, on peut dire qu’elle correspond à un enregistrement d’un spectrogramme.

L’étape de reconnaissance proprement dite consiste alors à analyser le signal inconnu sous la forme d’une suite de vecteurs acoustiques, et à comparer la suite inconnue à chacune des suites des exemples préalablement enregistrés. Le mot « reconnu » sera alors celui dont la suite de vecteurs acoustiques (le « spectrogramme ») colle le mieux à celle du mot inconnu.

Ce principe de base n’est cependant pas implémentable directement : un même mot peut en effet être prononcé d’une infinité de façons différentes, en changeant le rythme de l’élocution. Il en résulte des spectrogrammes plus ou moins distordus dans le temps. La superposition du spectrogramme inconnu aux spectrogrammes de base doit dès lors se faire en acceptant une certaine « élasticité » sur les spectrogrammes candidats. Cette notion d’élasticité est formalisée mathématiquement par un algorithme nommé : l’algorithme DTW (Dynamic Time Warping ou déformation dynamique temporelle).

On comprend donc qu’une telle technique soit limitée par la taille du vocabulaire à reconnaître (une centaine de mots tout au plus) et qu’elle soit plus propice à la reconnaissance mono-locuteur (une reconnaissance multi-locuteur imposerait d’enregistrer, de stocker, et surtout d’utiliser pour la comparaison de nombreux exemples pour chaque mot).

#### 4.2.2 Reconnaissance par modélisation d’unités de parole

La plupart des systèmes de reconnaissance de la parole vocale sont de nos jours basés sur ce mode là. Dès que l’on cherche à concevoir un système réellement multi-locuteur, à plus grand vocabulaire et s’adaptant facilement à une application, il devient nécessaire de mener la reconnaissance sur base d’unités de parole de plus petite taille, que l’on appelle phonèmes. En effet, la parole est constituée d'une suite de sons élémentaires. Ils sont produits par la vibration des cordes vocales. Ces sons mis bout à bout composent des mots. On ne se contente plus alors d’exemples de ces unités, mais on cherche plutôt à en déduire un modèle (un modèle par unité), qui sera applicable pour n’importe quelle voix. Il apparaît ainsi dans de nombreuses publications que l’on peut décomposer la reconnaissance de la parole en 4 modules :

* Un module d’acquisition et de modélisation du signal qui transforme le signal de parole en une séquence de vecteurs acoustiques. Pour être utilisable par un ordinateur, un signal doit tout d'abord être numérisé. Cette opération tend à transformer un phénomène temporel analogique, le signal sonore dans notre cas, en une suite d'éléments discrets.
* Un module acoustique qui peut produire une ou plusieurs hypothèses phonétiques pour chaque segment de parole (par exemple de 10 ms, pour chaque vecteur acoustique), associées en général à une probabilité. Ce générateur d'hypothèses locales est généralement basé sur des modèles statistiques de phonèmes, qui sont entraînés sur une grande quantité de données de parole (par exemple, enregistrement de nombreuses phrases) contenant plusieurs fois les différentes unités de parole dans plusieurs contextes différents.
* Un module lexical : Dans le cadre de la reconnaissance de la parole continue, même si le système acoustique est basé sur des phonèmes, il faut obtenir, pour chaque entrée du dictionnaire phonétique, un modèle qui lui est propre. Un tel module lexical embarque en général des modèles des mots de la langue (les modèles de base étant de simples dictionnaires phonétiques ; les plus complexes sont de véritables automates probabilistes, capables d’associer une probabilité à chaque prononciation possible d’un mot). A l’issue de ce module, il peut donc y avoir plusieurs hypothèses de mots qui ne pourront être départagées que par les contraintes syntaxiques.
* Un module syntaxique qui interagit avec un système d'alignement temporel pour forcer la reconnaissance à intégrer des contraintes syntaxiques, voire sémantiques. Les connaissances syntaxiques sont généralement formalisées dans un modèle de la langue, qui associe une probabilité à toute suite de mots présents dans le lexique. Ainsi le système est capable de choisir entre plusieurs mots selon le contexte de la phrase ou du texte en cours et de son modèle lexical.

On peut ajouter à cela un module de filtrage pouvant corriger le signal après l’acquisition afin de retirer les distorsions ou les bruits provenant du matériel ou de l’environnement du locuteur. Ce module est aussi appelé « traitement du canal de transmission ». Du fait de sa complexité et du peu d’amélioration qu’il apporte, ce module n’est pas toujours intégré aux systèmes. Cependant la recherche de meilleurs traitements du canal de transmission sera sûrement nécessaire à l’amélioration des systèmes de reconnaissance vocale.

## 5. Application

Cette partie va survoler les différentes possibilités d’appliquer la reconnaissance automatique de la parole.

### 5.1 Services vocaux

Les serveurs passifs (sans reconnaissance vocale) existent depuis de nombreuses années tels que l’horloge parlante, la météo, les résultats des courses, du loto, etc… Mais lorsque la quantité d’information est importante, il devient nécessaire pour l’utilisateur de pouvoir sélectionner ce qu’il veut entendre.

Dans des cas simples la sélection de touches « multifréquences » (DTMF) peut suffire. Mais pour des applications plus complexes des systèmes sont en voie de développement pour que l’utilisateur puisse naviguer sur un serveur vocal en prononçant les mots de contrôle de l’application. Ces services pourront s’étendre à tout un ensemble de domaines : la réservation de place d’avion, de train, de théâtre, de chambre d’hôtel, les déclarations de sinistre à l’assureur, les consultations et transactions bancaires, les opérations boursières, la facturation automatique des appels à distance, etc…

### 5.2 Contrôle de qualité, saisie des données

Dans de nombreux environnements de travail la possibilité de décharger le travailleur, grâce à une interface vocale, apporte un gain incontestable de liberté et de rapidité de mouvement. Pendant qu’il observe un processus complexe, il peut par exemple décrire des informations visuelles. Il a aussi la possibilité de commander à distance un automate évoluant en milieu hostile (apesanteur, sous-marin, industrie pétrolière).

### 5.3 Avionique

A bord des avions, les tâches étant de plus en plus complexes et le tableau de bord de plus en plus réduit, la parole permet au pilote d’avoir à sa disposition un moyen supplémentaire d’interaction avec la machine, sans cependant gêner l’accomplissement des tâches courantes qui requièrent de sa part toute son attention visuelle. Les autorités canadiennes ont été les précurseurs des techniques vocales dans l’avionique.

Ainsi, l’Institut de Recherche Aérospatiale (IRA) a effectué des travaux de recherche sur la technologie vocale depuis la fin des années 70.

Dans les années 80, la recherche sur la reconnaissance de la parole était axée sur la mise au point de techniques qui fonctionnent dans les postes de pilotage des aéronefs, où le niveau de bruit ambiant est très élevé. BAE Systems Canada a mis au point un système prototype de reconnaissance de la parole qui a été installé dans l’hélicoptère Bell 205 à stabilité variable de l’IRA. Ce système a démontré un très haut taux de reconnaissance de la parole à partir de données recueillies dans le poste de pilotage de l’hélicoptère, où il y a beaucoup de bruit.

### 5.4 Formation

Les enfants, mais aussi les adultes, sont attirés par des jeux doués de parole (poupées qui parlent, jeux de société, jeux vidéo, jeux éducatifs). L’enseignement assisté par ordinateur et notamment les laboratoires de langue commence à intégrer de plus en plus de possibilités vocales, et évolue vers une interactivité plus grande : les systèmes d’aide à l’apprentissage des langues étrangères, permettant d’acquérir une prononciation correcte, une maîtrise du vocabulaire et de la syntaxe, ne peuvent que bénéficier des technologies vocales qui leur confèrent en outre un aspect ludique.

### 5.6 Aide aux handicapés

Différents programmes européens ont permis de mieux cerner les différents types de handicap dont souffre la population, ainsi que le nombre de personnes concernées. On dénombre actuellement en Europe 12 millions de malvoyants dont 1 million de non-voyants, 81 millions de malentendants, dont 1 million de non-entendants, environ 30 millions de personnes ayant un handicap moteur des membres supérieurs et 50 millions ayant un handicap des membres inférieurs.

L’intérêt des technologies vocales apparaît évident dans la mesure où celles-ci permettent aux personnes handicapées de retrouver une certaine autonomie et de bénéficier d’une meilleure insertion dans leur environnement tant professionnel que familial, la parole se substituant au sens défaillant. Ainsi beaucoup de systèmes existent pour cela, tel que le contrôle de fauteuil roulant, le contrôle de fonctions secondaires dans la voiture, le contrôle d’appareil électrique à la maison, le contrôle de l’ordinateur, etc…

### 5.7 Autres applications

On peut aussi citer les modules de reconnaissance vocale embarqués, comme dans les téléphones mobiles ou les assistants numériques ainsi que les futures possibilités en cours de développement chez les fabriquants automobile avec le contrôle de différents éléments de la voiture grand public : autoradio, climatisation, navigation de bord. On peut même de nos jours surfer sur Internet grâce à des commandes vocales, c’est ce que propose la société InteractiveSpeech. L’utilisation de la reconnaissance automatique de la parole devient courante et devrait très bientôt apparaître dans la plupart des domaines d’activités et la plupart des applications futures.

## 6. Problématiques

L’environnement de la communication homme-machine est un espace interactif qui assure la communication entre l’utilisateur et la machine en différents modes (à travers le clavier, à travers les gestes, en utilisant les commandes vocales, etc.). De nos jours, comme nous avons précédemment mentionné, la communication vocale couvre pratiquement plusieurs domaines. Cependant, plusieurs facteurs peuvent empêcher les systèmes interactifs d’aboutir pour une communication performante, tels que :

* Facteurs internes : l’ambiguïté, l’implicite, modèle du langage, modèle acoustique.
* Facteurs externes : le bruit, le climat, le type de la parole (parole contenue, mots isolés).
* Facteurs dus aux locuteurs : une pathologie vocale, un locuteur non-natif, l’âge.

Une prononciation défectueuse peut falsifier l’information (commande vocale) voulue transmettre à la machine ce qui génère souvent des résultats de communication erronés. De plus, la progression de l’interaction homme-machine présente un manque remarquable en travaux de traitement automatique de la parole pathologique. Par conséquent, l’introduction des nouvelles techniques pour l’analyse des dégradations contenues dans un signal acoustique devient une exigence. Ces techniques permettent de corriger les anomalies qui peuvent toucher les commandes vocales à transmettre. D’ailleurs, la diversité des applications qui font intervenir l’interaction homme-machine, la sensibilité de certaines applications et le gain considérable apporté par ces applications en termes de temps et de productivité confirment la nécessité de vérification et d’analyse des inputs de ces applications. Cette analyse porte souvent sur le contenu de la commande vocale transmise. Parfois, ce contenu présente différentes dégradations dues aux problèmes articulatoires.

## 7. Difficultés d’ordre pathologique

Un locuteur qui souffre d’une pathologie vocale produit souvent un énoncé défectueux. Cet énoncé peut falsifier l’information (commande vocale) voulue transmettre à la machine ce qui génère souvent des résultats de communication erronés. D’autre part, la littérature présente un manque remarquable en travaux de traitement automatique de la parole pathologique. De ce fait, l’introduction des nouvelles techniques pour l’analyse des dégradations contenues dans les sons voisés devient une exigence. Ces techniques permettent de corriger les anomalies qui peuvent toucher les commandes vocales transmises. De même, la diversité des applications qui font intervenir l’interaction homme-machine, la sensibilité de certaines applications et le gain considérable apporté par ces applications en termes du temps et de productivité confirment la nécessité de la vérification et de l’analyse des inputs pour ces applications.

Selon les paramètres qui caractérisent la voix humaine, la parole produite peut suivre plusieurs classifications possibles. Mais, comme l’objectif de ce travail consiste à détecter la parole pathologique, nous nous limitons volontairement à la classification en parole saine (normale) ou parole pathologique. Nous citons pour chaque classe les traits qui la caractérisent.

### 7.1 Parole saine

Une parole saine est celle produite par un locuteur qui maîtrise et respecte toutes les règles de prononciation de la langue du discours. Une parole saine ne contient pas aucune anomalie dans ces constitutions. C’est donc la parole qui n’est pas atteinte par aucune déformation. Généralement, une parole saine est caractérisée par :

* Bonne articulation : chaque phonème est couvert par sa propre zone articulatoire.
* Fréquence et intensité acceptables : un compromis entre le nombre et l’amplitude des vibrations des cordes vocales.
* Bonne distribution de l’énergie sonore : la vitesse de production de la parole facilite la compréhension pour la partie réceptrice (homme ou machine).

### 7.2 Parole pathologique

La voix humaine peut être atteinte par de nombreux dysfonctionnements tels que les défauts de prononciation dus aux lésions des cordes vocales. De cette façon, la parole pathologique est celle qui est partiellement ou totalement défectueuse. Autrement dit, c’est une suite de la parole qui contient des déformations au niveau de ses composantes fondamentales qui sont les phonèmes.

Une parole pathologique engendre généralement des problèmes de compréhension pour les personnes réceptrices. Ce problème augmente proportionnellement avec le taux de fausses prononciations contenues dans la parole produite et s’approfondit dans le cas de la communication homme-machine.

Les déformations qui peuvent toucher une parole humaine sont souvent de genre :

* Lésions des cordes vocales : ils empêchent les cordes vocales d’atteindre la fréquence ou l’intensité qui correspondent à la pression d’air sous-glottique.
* Mauvaise articulation : un ou plusieurs phonèmes ne se produisent pas de leurs propres points articulatoires.
* Paramètres prosodiques non maîtrisés (vitesse d’élocution trop élevée/trop faible, mauvaise distribution de l’énergie acoustique, etc.).

### 7.3 Typologie des pathologies vocales

Les pathologies vocales suivent plusieurs types, vu la diversité des facteurs qui les posent. Nous présentons dans cette section les différents types de pathologies vocales dues à des problèmes articulatoires : la suppression, la substitution, la distorsion, la permutation et l’addition.

1. Suppression

Ce type de défauts de prononciation peut être exprimé par la suppression d’un ou plusieurs morphèmes de la prononciation désirée. Par exemple, le locuteur peut produire la fausse prononciation « كُوة » à la place de celle désirée « كُورة » en supprimant le morphème « ر ». Dans ce cas de troubles de langage, la difficulté de compréhension de la parole produite augmente si la suppression dépasse un morphème.

1. Substitution

Dans ce type de troubles de prononciation, le locuteur remplace un ou plusieurs sons désirés par d’autres sons. Par exemple, il produit la prononciation « مدرثة » à la place de celle « مدرسة » en remplaçant le morphème « س » par celui « ث ». Ce type des pathologies vocales est plus connu chez les enfants que dans les discours des adultes. Dans un tel cas, la compréhension d’un discours produit devient plus difficile si les substitutions sont fréquentes.

1. Distorsion

Il s’agit d’une manifestation de la substitution. Ce type de pathologies vocales apparaît lorsque le son prévu est mal prononcé, mais le nouveau son est proche de celui désiré. Par exemple, le locuteur produit la fausse prononciation « تاولة » à la place de celle désirée « طاولة » en remplaçant le morphème « ط » par celui « ت ». Cette pathologie vocale est plus connue dans les discours des adultes que chez les enfants et les jeunes. Bien qu’une telle pathologie soit produite, la nouvelle prononciation reste compréhensible dans les communications homme à homme mais elle pose des problèmes de compréhension dans les communications homme-machine.

1. Permutation

Cette pathologie vocale est très connue chez les enfants. Elle se manifeste si le locuteur permute entre deux phonèmes de la prononciation attendue. Par exemple, les locuteurs peuvent produisent la fausse prononciation « السّمش » à la place de celle désirée « الشّمس » en permutant les deux phonèmes « ش » et « س ». Dans ce type de défauts de prononciation, on ne parle pas de phonèmes problématiques comme seulement l’ordre des phonèmes qui n’est pas maîtrisé.

1. Addition

C’est la pathologie la moins connue. Elle apparaît lorsque le locuteur ajoute un ou plusieurs sons à la prononciation prévue. L’exemple le plus connu est l’ajout du son « ن » ou « م » à toutes les prononciations qui se terminent par un « Mâd/ مدّ » (« ا ou ى », « ي » et « و »).

### 7.4 Facteurs sources des pathologies vocales

Dans cette section nous dressons les principaux facteurs qui posent des troubles vocaux engendrant des difficultés d’élocution. Ces défauts de prononciation peuvent être d’origine :

* Apprenant non-natif

L’apprentissage des langues étrangères engendre une diversité d’accents chez les locuteurs non-natifs. Souvent l’accent d’un tel apprenant présente plusieurs déformations dans la parole produite puisqu’il est gêné par le contexte articulatoire de sa langue maternelle.

* Etat psychologique

La situation sentimentale du locuteur influe soit positivement soit négativement sur la production de la parole. En effet, un locuteur dans un état d’âme défavorable risque de produire une parole pathologique. Par exemple, probablement, la parole produite par un locuteur qui pleure va contenir des déformations fondamentales.

* Facteur anatomique

Dans ce cas, les locuteurs présentent des malformations de l’appareil buccophonatoire, telles que des fentes labio-palatines. Généralement, la surdité profonde s’accompagne également de troubles de production de la parole [3].

* Facteur neurologique

Les locuteurs qui souffrent de troubles de prononciation d’origine neurologique ont généralement une faiblesse ou un manque de coordination dans les mouvements articulatoires [4]. Cette faiblesse est engendrée par des atteintes du système nerveux comme la maladie de Parkinson ou la sclérose en plaques.

* Environnement acoustique

L’environnement acoustique a un impact sur la parole produite par un locuteur, et également sur la perception en termes de ressources mentales allouées par l’auditeur pour traiter le message transmis. Cet impact de l’environnement acoustique peut être soit positif, soit négatif. Il est reconnu qu’un environnement acoustique défavorable est un facteur fondamental de risque pour la dégradation de la voix [5].

## 8. Conclusion

On peut dire que depuis une décennie, les techniques de traitement de la parole ont connu plusieurs grandes révolutions.

La première, et celle qui touche pour l’instant de loin le plus d’utilisateurs, est celle de la téléphonie mobile : une proportion grandissante de la population transporte avec elle un ordinateur de poche avec un module de reconnaissance vocale. Les algorithmes de codage sont par ailleurs également utilisés dans les boîtes vocales : nos paroles y sont stockées sous la forme de suites de vecteurs de paramètres. Le marché du codage de la parole est donc à présent largement ouvert.

La seconde révolution est celle des grandes bases de données de parole et de textes. Depuis 1995, aux Etats-Unis et en Europe, de nombreux laboratoires de recherche (publics et privés) mettent en commun leurs ressources. Il en résulte un foisonnement de données propices à l’établissement de modèles, tant numériques que symboliques, de la parole. Les développements récents en reconnaissance, et plus encore en synthèse, en sont en grande partie la conséquence logique.

Enfin, une dernière révolution se prépare : celle qui verra naître des machines dont plus personne ne pourra affirmer avec certitude qu’elles en sont. Aujourd’hui déjà, la qualité des algorithmes de synthèse vocale permet aux synthétiseurs de passer avec succès le « test de Turing », inventé par le mathématicien anglais Alan Turing dans les années 40 pour mesurer le degré d’« intelligence » d’une machine : en vérifiant combien de temps un expérimentateur interagissant « en aveugle » avec cette machine peut rester persuadé d’avoir affaire à un être humain. La reconnaissance est aussi prête à tromper notre intelligence, et ne manquera pas de continuer à le faire, dès lors que l’on aura amélioré ses capacités de robustesse.

Devant la diversité des facteurs posant des problèmes de prononciation pour les locuteurs de l’arabe d’un côté, les limites des interventions cliniques pour corriger les déformations du système phonatoire engendrant des défauts d’élocution de l’autre côté, l’intervention des outils de traitement automatique de la parole, des techniques linguistiques et des approches formelles peuvent faciliter énormément la tâche de détection et de correction automatique de la parole pathologique. Dans le chapitre suivant, nous détaillerons la méthode utilisée dans ce travail pour la détection des pathologies vocales dans le discours Arabe.

Il n’en reste pas moins que, alors que nos ordinateurs pourront nous parler et reconnaître ce que nous leur dirons, ils n’en seront pas pour autant capables de comprendre nos paroles. C’est là un tout autre domaine, dont nous ne connaissons encore que les tous premiers balbutiements.

# Chapitre 2 : Choix Méthodologique et Environnement Technique

## 1. Introduction

Le présent chapitre fait l’objet de la description de l’approche proposée pour la détection des pathologies vocales contenues dans la parole arabe et aussi de l’environnement du travail.

## I. La méthode de détection des pathologies vocales

## 1. Introduction

L’approche proposée pour la détection des pathologies vocales contenues dans la parole arabe se base, d’une part, sur la modélisation phonétique de la parole produite et d’autre part, sur la distance angulaire entre le modèle phonétique référence de l’arabe parlée et celui propre au locuteur, comme mesure de similarité.

Deux principales raisons peuvent justifier le choix de l’approche probabiliste basée sur la modélisation phonétique :

* La représentation de la parole produite sous forme d’un modèle phonétique (vecteur des probabilités) est une représentation globale qui n’engendre pas une perte de données. Donc, on traite l’information en totalité et non pas un extrait local qui parfois ne reflète pas le vrai contenu de l’énoncé.
* Les avantages des approches numériques dont les données sont plus faciles à manipuler : en effet, la conversion du signal voisé en un vecteur des probabilités, facilite la comparaison de la parole produite à celle référence énormément que travailler avec la parole elle-même.

Le signal de parole produit par un locuteur peut contenir des caractéristiques qui le distinguent d'un autre signal : parole pathologique ou saine. La méthode à suivre est axée sur la détection des prononciations faussées dans les discours arabes. La transcription phonétique peut être utilisée pour générer un modèle phonétique de la parole arabe : le pourcentage d'occurrences de chaque bi-phonème dans la langue arabe parlée. La comparaison entre le modèle phonétique standard du discours arabe et celui propre au locuteur permet de classer le discours produit par le locuteur concerné en pathologique ou en bonne santé.

La première étape de l’approche consiste à générer une distance standard (distance phonétique) entre différentes bases de langage arabe saine. Dans la deuxième étape, nous déterminerons la distance entre le discours à classer et le discours de référence (modèle phonétique propre au locuteur et modèle phonétique de référence). En comparant ces deux distances (distance entre parole à classer et référence paroles et distance standard), dans la troisième étape, on peut classer la parole entrée en saine ou pathologique.

## 2. Génération d’un modèle phonétique

### 2.1 Préparation des bases de langage arabe

Pour avoir des bases vocales (corpus) arabe saines, soit on enregistre des gens qui parlent pendant des heures et ensuite on génère la transcription, en utilisant un logiciel (ou service) de reconnaissance automatique de la parole, soit on cherche des corpus sains prêts. Nous avons adopté la deuxième solution.

Quatre corpus ont été choisis :

* Le saint Coran

Des articles des journaux dans trois domaines différents :

* Sport
* Culture
* International
* Politique
* Economie

### 2.2 Nettoyage des corpus

La phase de nettoyage des corpus consiste à :

* Conserver que les lettres arabes.
* Enlever les caractères spéciaux (ponctuations, numéros, …).
* Enlever les signes diacritiques.
* Normalisation des lettres.

### 2.3 Modélisation phonétique

En utilisant les fichiers contenant les transcriptions, on calcule les probabilités d’occurrence de chaque bi-phonème possible dans le discours Arabe. L’arrangement de toutes les probabilités dans un vecteur de 784 coefficients forme le modèle phonétique référence pour la parole Arabe (784=282 : la langue Arabe compte 28 phonèmes.)

## 3. Approche proposée

La méthode proposée pour la détection automatique de la parole Arabe pathologique est une méthode qui se base sur la comparaison entre les modèles phonétiques référence de la parole Arabe et celui propre au locuteur pour détecter la pathologie. Cette comparaison utilise la distance angulaire entre les deux modèles phonétiques comme mesure de similarité.

La stratégie de détection proposée repose sur quatre principales phases. La première concerne la génération de n modèles phonétiques de la parole arabe (en utilisant un corpus pour chaque modèle phonétique) afin de calculer la distance maximale entre les modèles phonétiques arabes (distance phonétique). La deuxième phase s’intéresse à la génération du modèle de référence phonétique (moyenne de n modèles précédents). En troisième lieu, pour chaque nouvelle séquence à classer, nous générons le modèle phonétique propre au locuteur (le locuteur peut être normal, natif, avec un handicap…). Enfin, comparez ces deux modèles et classifiez le discours en entrée saine ou pathologique.

### 3.1 Détection des pathologies vocales

#### 3.1.1 Produit scalaire dans

On appelle produit scalaire dans et on note la fonction qui associe aux vecteurset la quantité :

Cette quantité peut être exprimée aussi comme suit :

Avec :

* est l’angle formée par les deux vecteurs et .
* correspond à la norme du vecteur ,
* correspond à la norme du vecteur ,

On déduit alors que :

En représentant les fichiers audios sous forme de vecteurs dans , on peut utiliser l’angle comme mesure de similarité entre les fichiers représentés par ces vecteurs. En effet, plus la similarité est forte entre les signaux acoustiques, plus la distance angulaire (l’angle ) entre les vecteurs représentatifs est faible.

#### 3.1.2 Distance phonétique et classification de la parole

La génération de la distance phonétique nécessite :

* La préparation de n sous-corpus de la parole Arabe , et pour chaque souscorpus , nous déterminons le modèle phonétique qui lui correspond.
* La définition de l’ensemble des angles qui séparent les différents modèles et ( est l’angle qui sépare le modèle de , ).
* La définition de la valeur
* La définition de la valeur
* La génération de la valeur
* Le calcul de , où est une marge de tolérance.

Pour calculer les angles qui forment l’ensemble S, nous suivons les formules du produit scalaire (1), (2) et (3) définies dans la section précédente.

Pour chaque nouveau locuteur, on utilise une séquence enregistrée par sa voix et on suit la même procédure précédente pour générer son propre modèle phonétique.

Nous calculons par la suite l’angle α qui sépare ce modèle de celui référence de la parole Arabe. On distingue deux cas possibles :

* Si , la parole en entrée (prononcée par le locuteur concerné) est saine ; donc aucun problème de prononciation.
* Sinon la parole en entrée est pathologique.

## 4. Récapitulation

La méthode proposée pour classer la langue arabe peut être résumée comme suit :

* Génération de modèles phonétiques de la parole arabe (en utilisant un corpus pour chaque modèle phonétique) afin de calculer la distance maximale entre les modèles phonétiques arabes (distance phonétique).
* Génération du modèle de référence phonétique (moyenne de n modèles précédents).
* Pour chaque nouvelle séquence à classer, nous générons le modèle phonétique propre au locuteur (le locuteur peut être normal, natif, avec un handicap…).
* Comparer ces deux modèles et classifier le discours en entrée saine ou pathologique.

## II. Environnement technique

## 1. Python

Python est un langage de programmation open source, libre et multi-plateformes, à la fois puissant et facile à maîtriser. Il est largement utilisé et pris en charge. Pour en savoir plus sur Python, reportez-vous au site Web python.org.

J’ai utilisé python pour l’implémentaion de la méhtode de détection des pathologies, en tant que service suivant l’architecture REST.

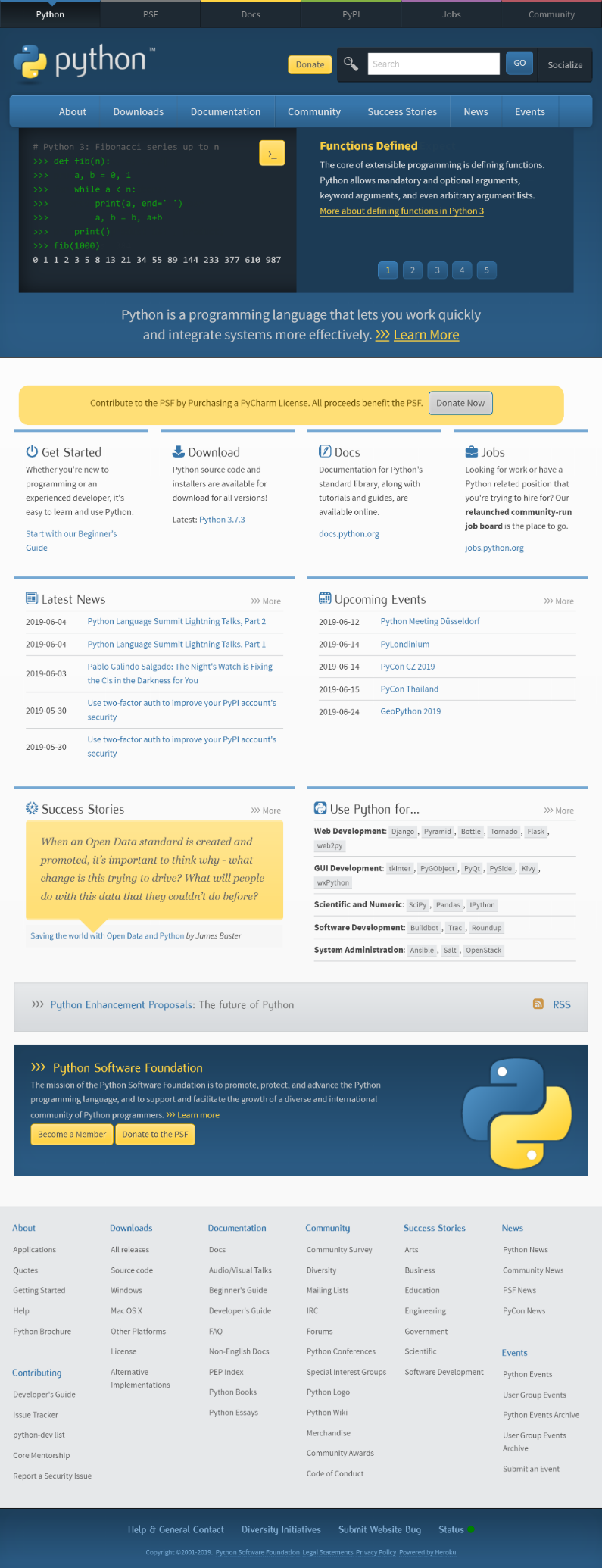


Figure 1 - Site web de Python

## 2. NumPy

NumPy est une extension du langage de programmation Python, destinée à manipuler des matrices ou tableaux multidimensionnels ainsi que des fonctions mathématiques opérant sur ces tableaux.

## 3. Flask

Flask est un micro-framework open-source de développement web en Python. Son but principal est d'être léger, afin de garder la souplesse de la programmation Python, associé à un système de templates. Il est distribué sous licence BSD.

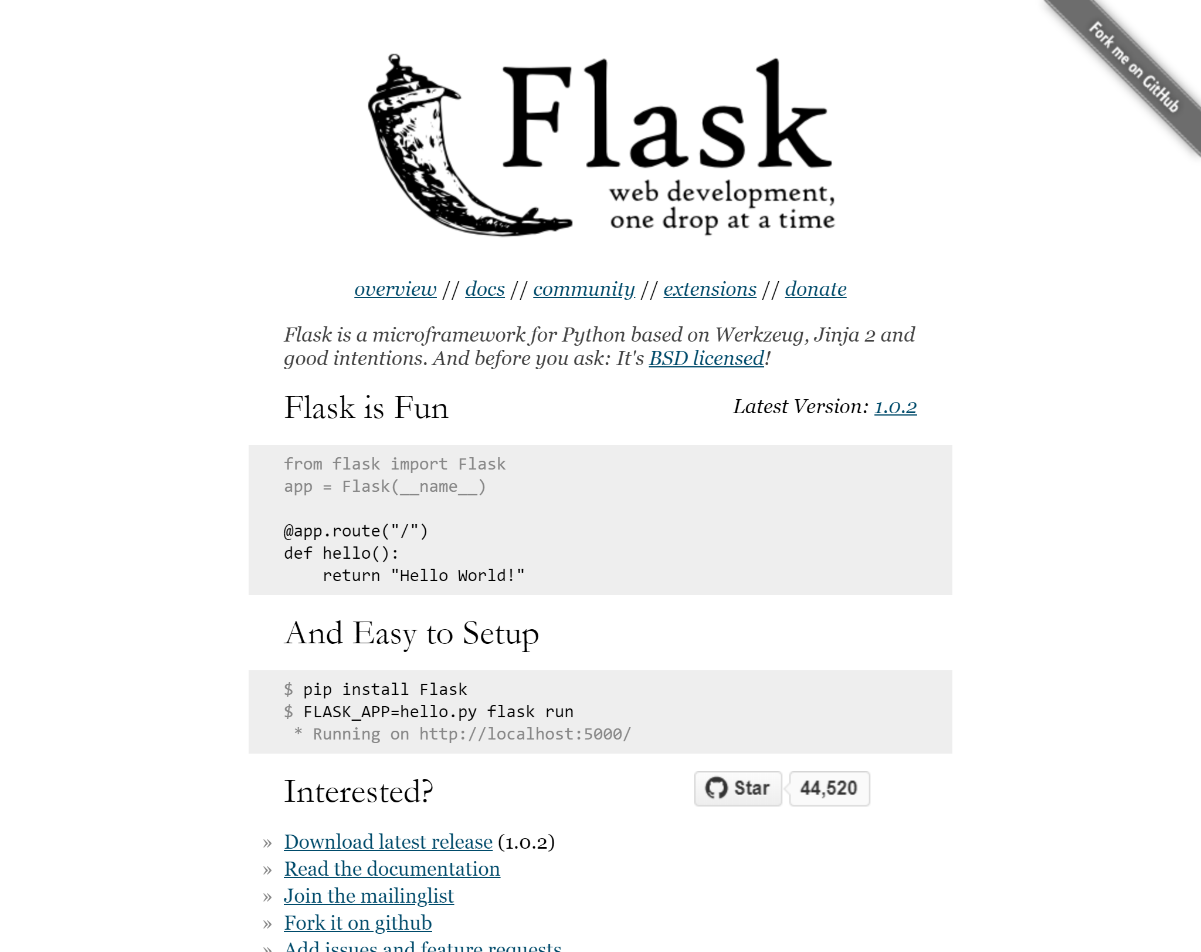


Figure 2 - Site Web Flask "http://flask.pocoo.org/"

## 4. Flask-RESTful

Flask-RESTful est une extension pour Flask qui prend en charge la création rapide d’API REST.



Figure 3 - Site Web FalskRESTful <https://flask-restful.readthedocs.io/en/latest/>

## 5. Java

Java est un langage de programmation qui permet de développer des applications informatiques pour différents systèmes d’exploitation. Il s’agit d’un langage orienté objet conçu par James Gosling et Patrick Naughton, qui est apparu en 1995 et très populaire dans l’informatique professionnelle. Il est aujourd’hui maintenu par la société Oracle.

## 6. Android

Android est un système d'exploitation mobile développé par Google. Il repose sur une version modifiée du noyau Linux et d'autres logiciels à code source ouvert, et est principalement conçu pour les appareils mobiles à écran tactile tels que les smartphones et les tablettes.

## 7. PocketSphinx

PocketSphinx est un moteur de reconnaissance vocale léger, spécialement conçu pour les appareils de poche et mobiles, bien qu'il fonctionne aussi bien sur le bureau.

## 8. PyCharm

PyCharm est un environnement de développement intégré utilisé pour programmer en Python.

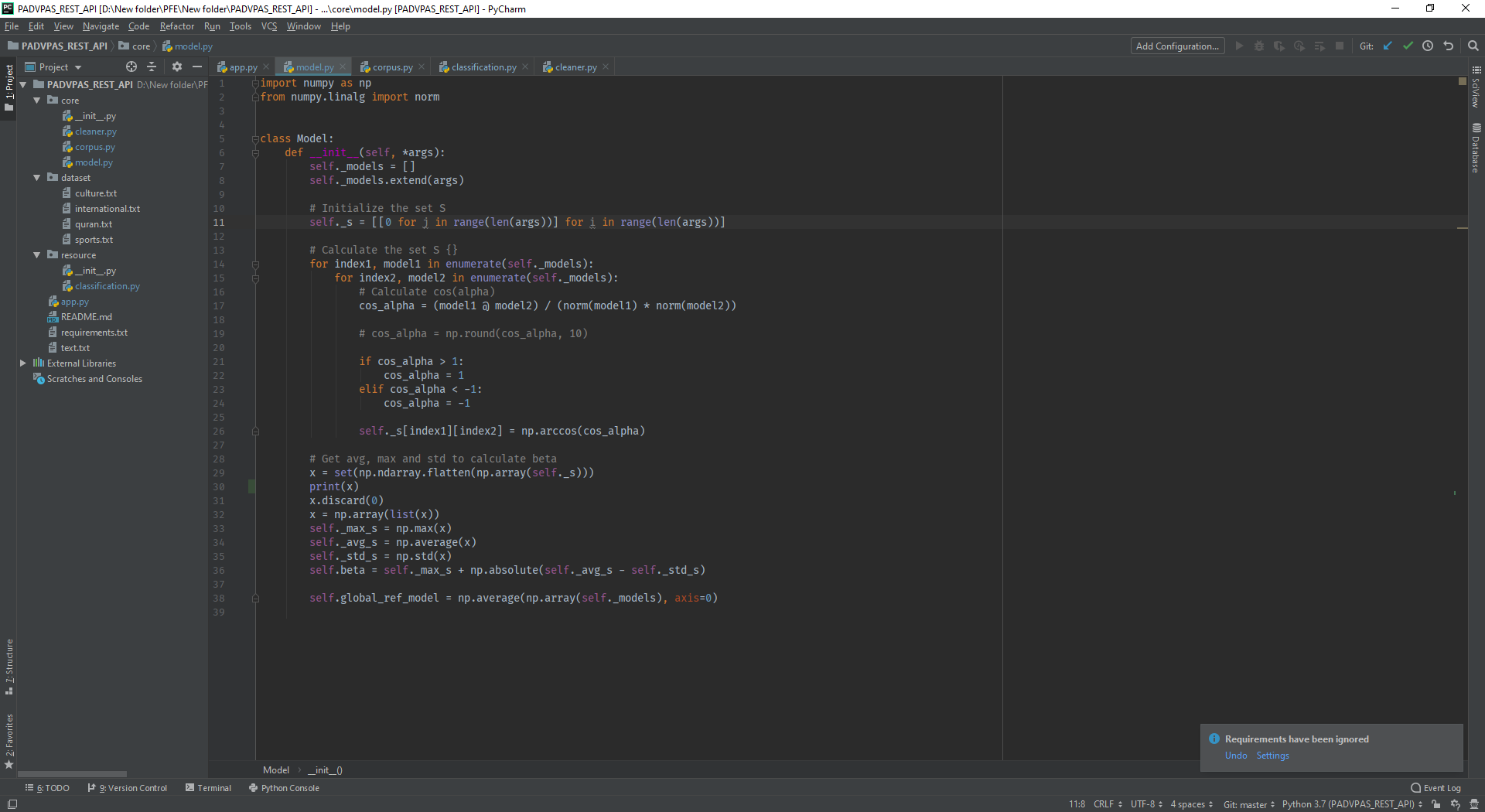


Figure 4 - PyCharm

## 9. Android Studio

Android Studio est l'environnement de développement intégré officiel du système d'exploitation Android de Google, construit sur le logiciel IntelliJ IDEA de JetBrains et conçu spécifiquement pour le développement Android.

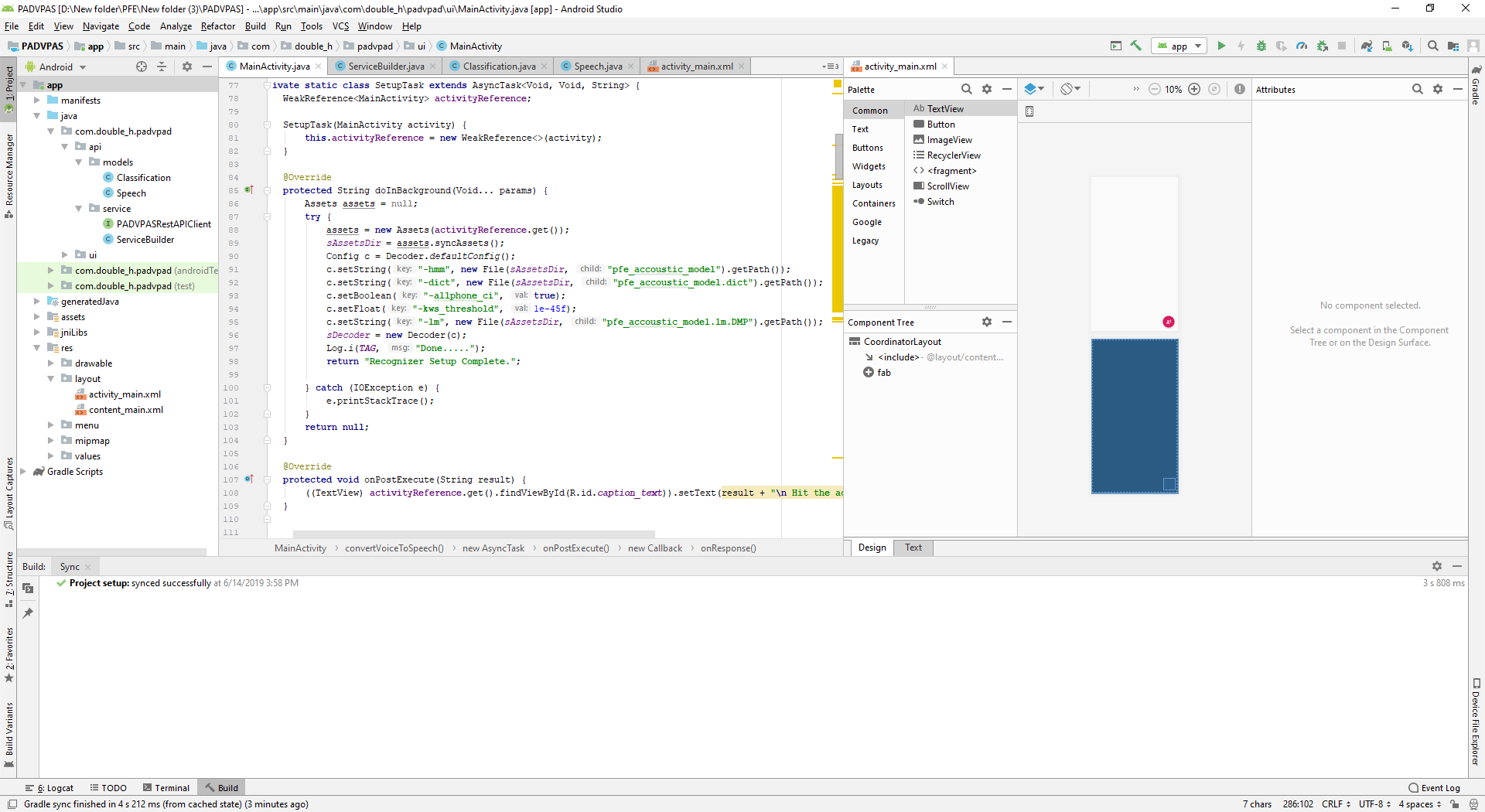


Figure 5 - Android Studio

## 10. Postman

Postman est un excellent outil pour tenter de disséquer les API RESTful. Il offre une interface utilisateur élégante avec laquelle faire des requêtes HTML, sans avoir à écrire un tas de code juste pour tester les fonctionnalités d'une API.

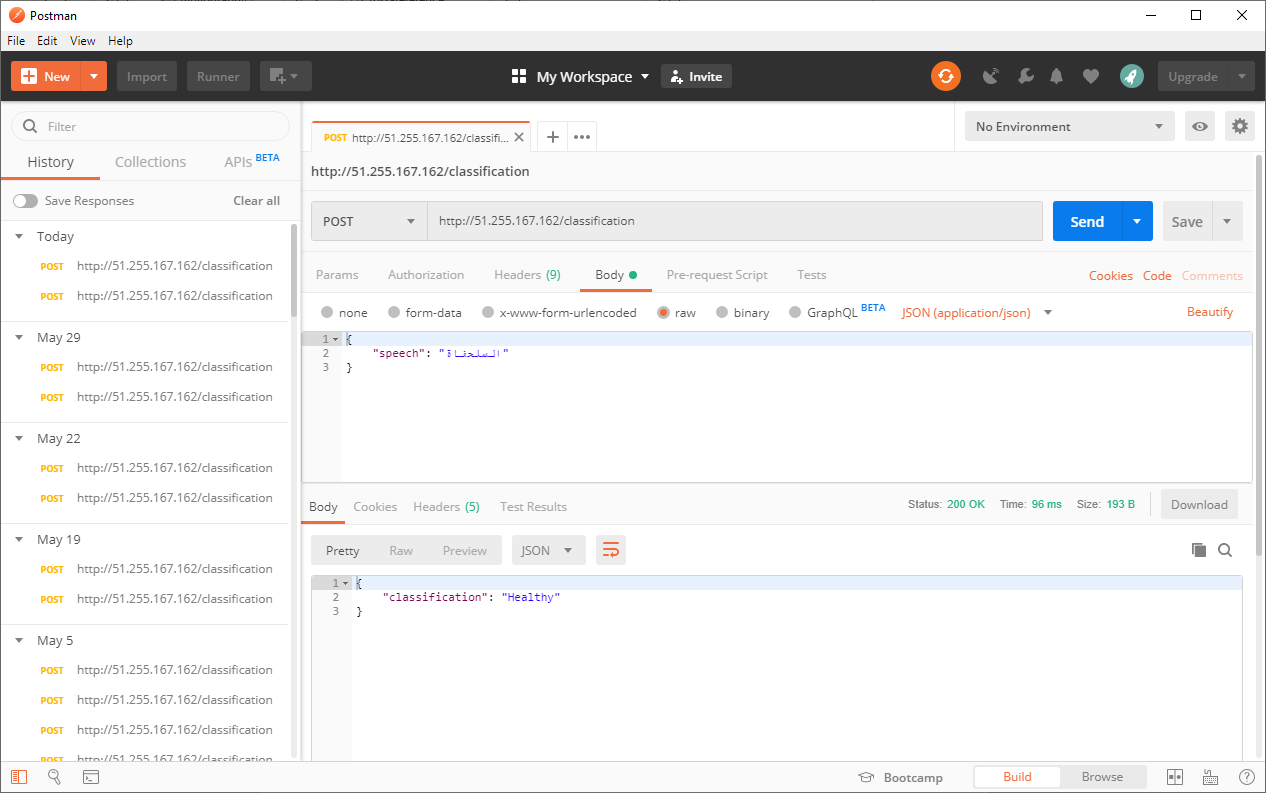


Figure 6 - Postman

## 11. Visual Studio

Microsoft Visual Studio est un environnement de développement intégré de Microsoft. Il est utilisé pour développer des programmes informatiques, ainsi que des sites Web, des applications Web, des services Web et des applications mobiles.

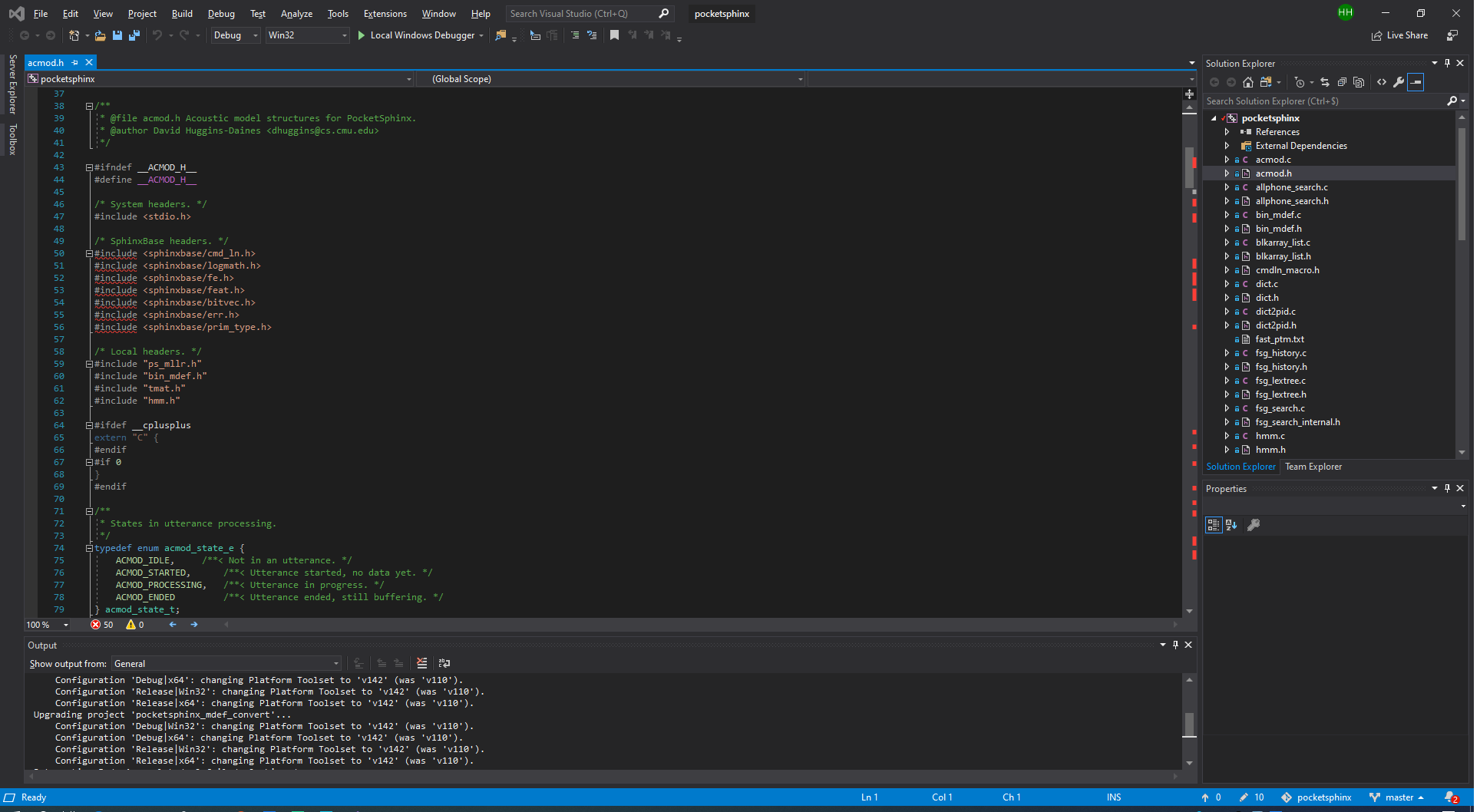


Figure 7 - Visual Studio

## III. Conclusion

Dans ce deuxième chapitre, nous avons, décrit la méthode utilisée pour la détection des pathologies vocales contenues dans le discours arabe, cité les différentes technologies et environnements de travail utilisées qui ont rendues facile la réalisation de ce travail.

# Chapitre 3 : Réalisation

## 1. Introduction

Ce projet intitulé « Application mobile pour la détection des pathologies vocales dans le discours Arabe » est divisé en deux grandes tâches, un service de détection des pathologies, et une application mobile servant comme une interface.

* Application mobile :
  + Une interface
  + La reconnaissance automatique de la parole se fait au niveau de l’appareil mobile.
* Service de détection des pathologies :
  + Implémentation de la méthode décrite dans le chapitre 2.

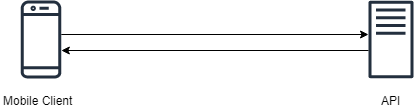


Figure 8 – Description abstraite de la communication entre l’application mobile et l’API REST

## 2. Le service de détection des pathologies vocales dans la parole Arabe

### 2.1 Les services web

Les services web sont des protocoles qui permettent la communication et l’échange de données entre des systèmes informatiques distribués. Le grand avantage des services web c’est que la communication peut se faire entre des systèmes informatiques avec des architectures différentes. Les systèmes qui exploitent les services peuvent être dans des environnements de développement totalement différents.

#### 2.1.1 Types des services web

Il existe deux principaux types de service web :

* Les services web WS

Il s’agit des services web basés sur les standards « SOAP » et « WSDL ». Ces services exposent les fonctionnalités sous la forme de services exécutables à distance. Les services sont ensuite répertoriés dans des annuaires comme les « UDDI » (Universal Description, Discovery, and Integration).

* Les servies web de type REST

Les services web de type « REST » (representational state transfer) appelé aussi « RESTful Web Services » sont des services web basés sur les technologies du web. C’est-à-dire qu’ils exposent leurs fonctionnalités à travers des « URI » accessibles par le protocole « HTTP ». Les messages de réponse peuvent avoir des représentations différentes (Texte, XML, JSON, …).

Notre implémentation du service est suivante l’architecture RESTfull.

### 3. Déploiement du service

Dans cette section je vais citer les étapes pour déployer le service de détection des pathologies vocale contenu dans le discours Arabe à un serveur « VPS : Virtual Private Server ».

Le déploiement du service consiste à configurer le VPS (Vérification des mis à jour, création d’un nouvel utilisateur, activation du pare-feu, etc.), ensuite l’installation des packages nécessaire au bon fonctionnement du service (nginx, python, …), et l’étape finale sera le test du fonctionnement du service avec l’outil de test des API REST Postman.

Remarque : La/les commande(s) pour exécuter les tâches suivantes seront toute(s) débutées avec le caractère $. Une explication de la tâche serait incluse si nécessaire.

* Accès au server

Avec la commande suivante on accède à notre VPS, après avoir fournir le mot de passe de l’utilisateur root (Les données de connexion au VPS sont obtenu à partir de l’hébergeur après avoir payé leur service « le VPS »)

$ ssh [root@vps677061.ovh.net](mailto:root@vps677061.ovh.net)

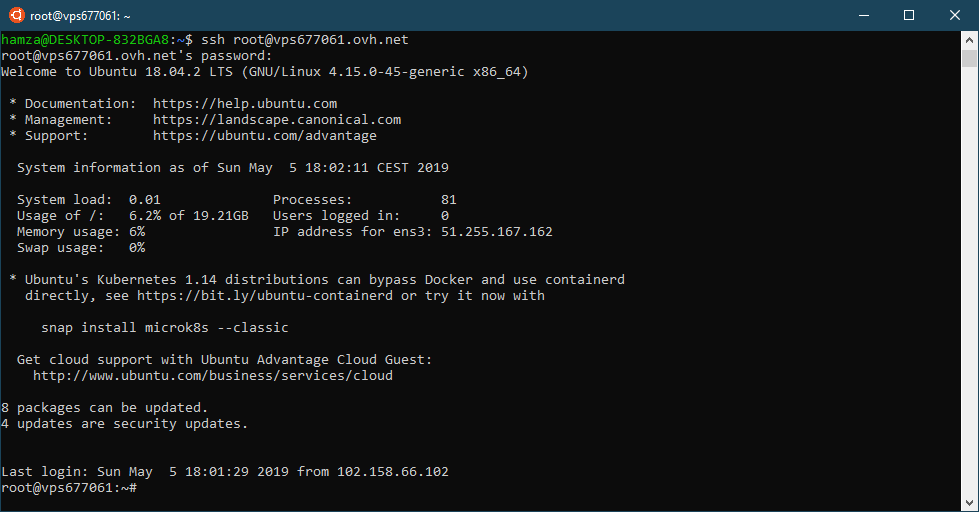


Figure 9 - Premiers accès au serveur

* Vérification/ Installation des mis à jour

Une vérification des mis à jour est faite avec les deux commandes suivantes, la première pour chercher les mis à jour, la dernière pour les téléchargées et les installées.

$ apt update

$ apt upgrade

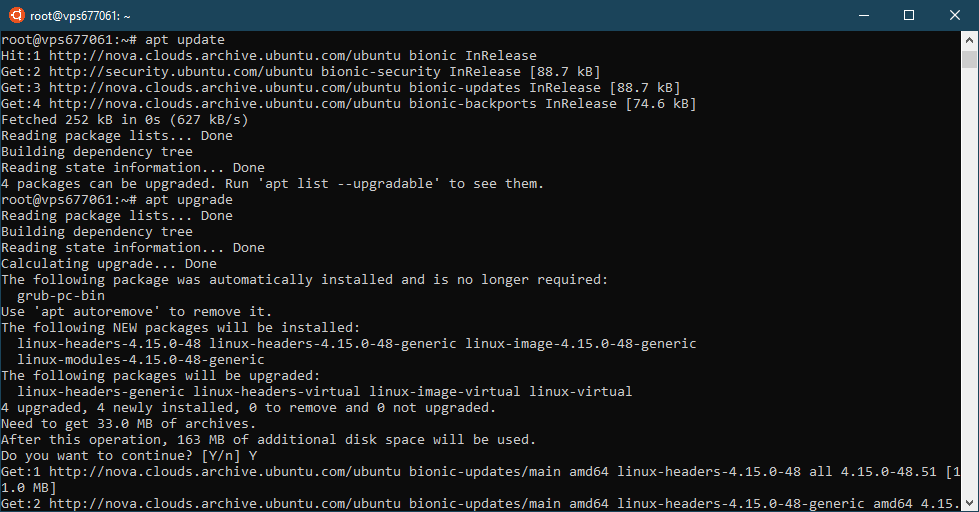


Figure 10 - Vérification des mis à jour

* Ajout d’un nouvel utilisateur

Une bonne pratique exige la création d’un nouvel utilisateur, et de continuer la configuration du VPS en tant que ce dernier à la place de l’utilisateur root.

$ adduser hamza

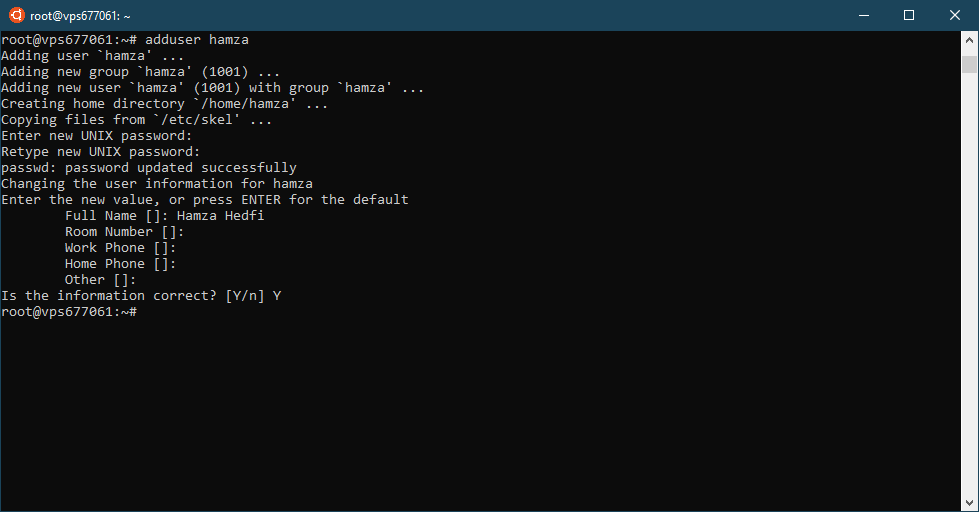


Figure 11 - Ajout d'un nouvel utilisateur

Maintenant on doit accorder au nouvel utilisateur les pouvoir de l’utilisateur root ; C’est en ajoutant au fichier sudoers.tmp situé à /etc/ la ligne suivant

hamza ALL=(ALL:ALL) ALL

$ nano /etc/sudoers.tmp

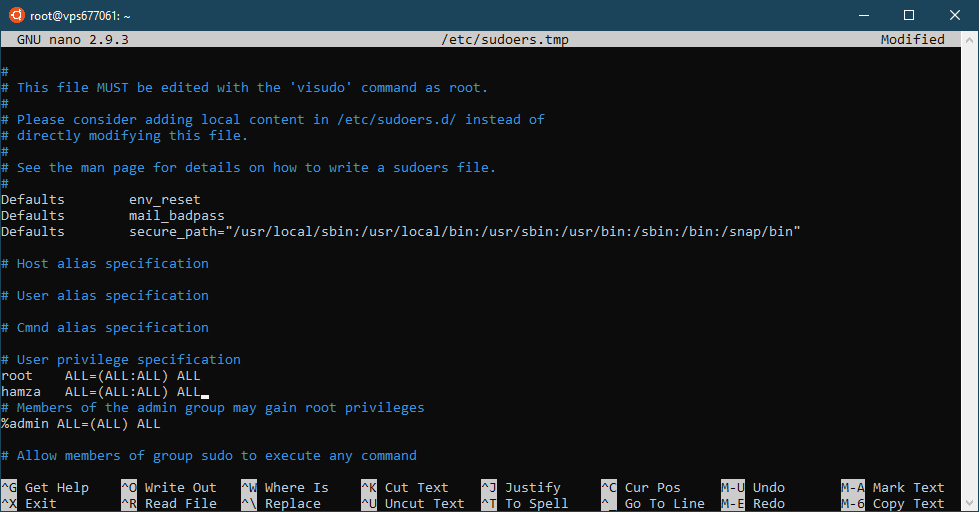


Figure 12 - Donner au nouvel utilisateur les privilèges root

* Bloquer l’accès de l’utilisateur « root » à distance et en revanche permettre l’accès au nouvel utilisateur

Pour mieux garantir la sécurité de notre VPS, on doit bloquer l’accès à distance de l’utilisateur root, ensuite on doit permettre au nouvel utilisateur « hamza » l’accès à distance. Alors on doit modifier le fichier sshd\_config située à /etc/ssh/ en modifiant la ligne PermetRootLogin yes à PermetRootLogin no et enfin on ajoute la linge AllowUsers hamza

$ vim /etc/ssh/sshd\_config

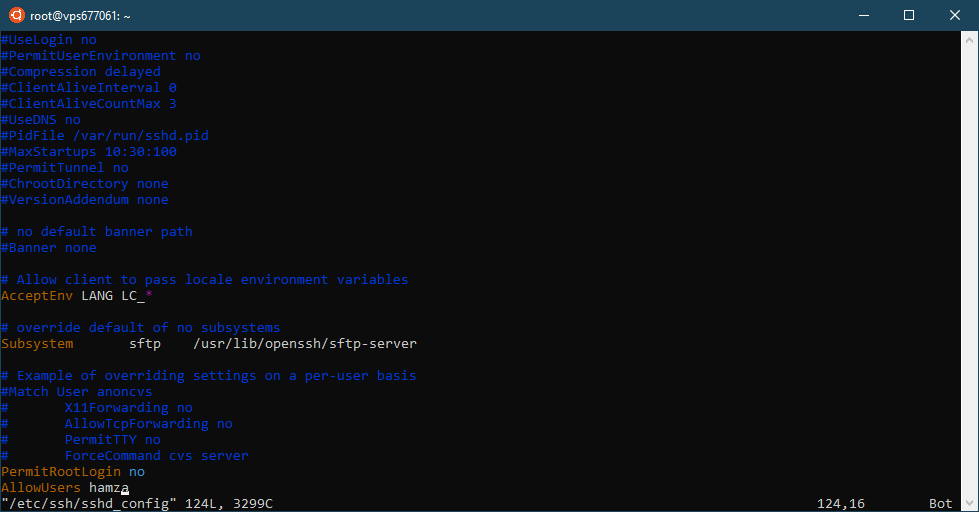


Figure 13 - Modification du service ssh

A la fin on redémarre le service ssh pour appliquer les nouvelles modifications.

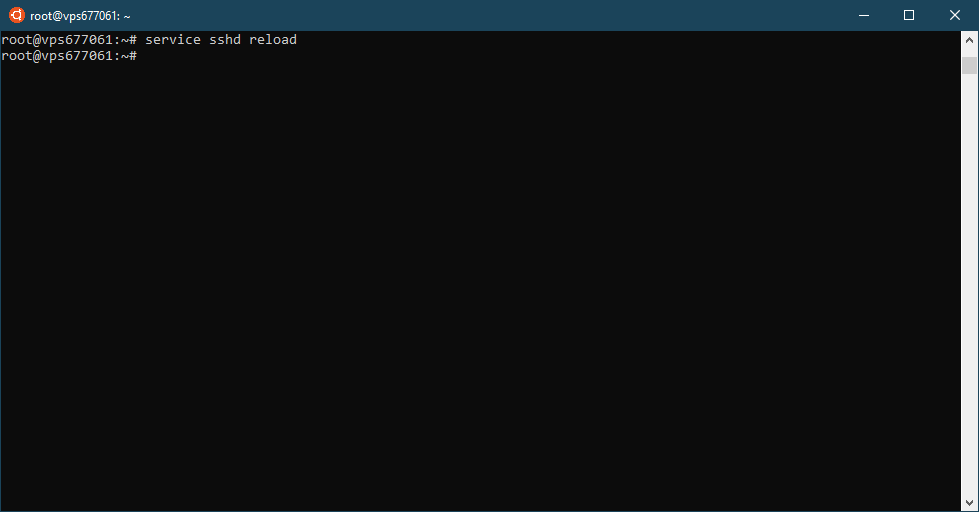


Figure 14 - Redémarrage du service ssh

Après les dernières modifications, on se reconnecte au serveur on tant que le nouvel utilisateur.

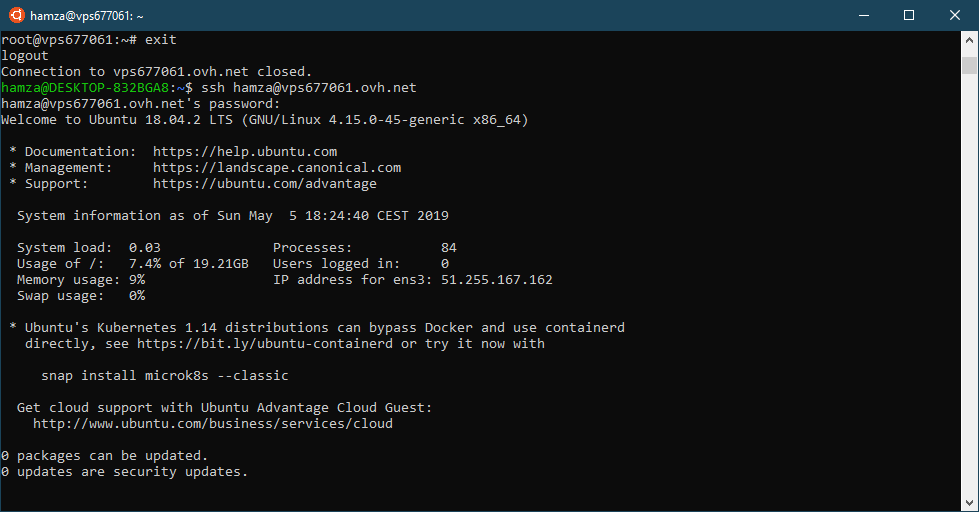


Figure 15 - Reconnecter au server en tant qu’utilisateur hamza

* Installation de nginx [[1]](#footnote-1)

$ sudo apt install nginx

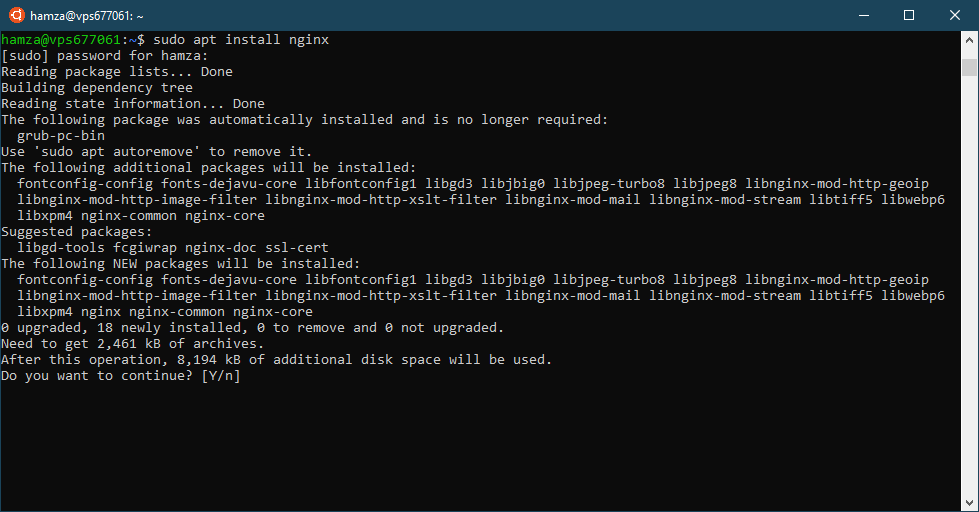


Figure 16 - Installation de nginx

* Activation du pare-feu

Pour assurer une meilleure sécurité du serveur on active le pare-feu « UFW[[2]](#footnote-2) ».

$ sudo ufw enable

On assure ensuite les communication HTTP et SSH à travers le pare-feu.

$ sudo ufw allow ‘Nginx HTTP’

$ sudo ufw allow ssh

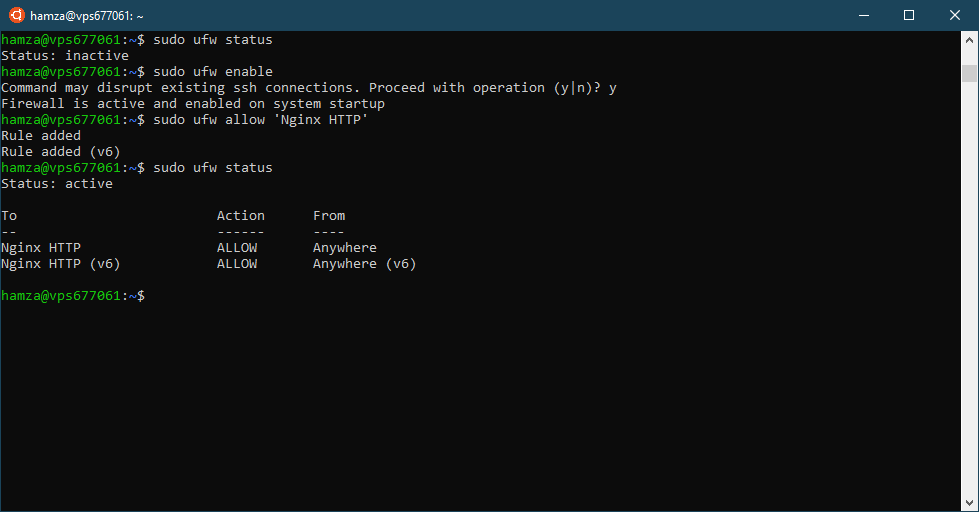


Figure 17 - État du pare-feu

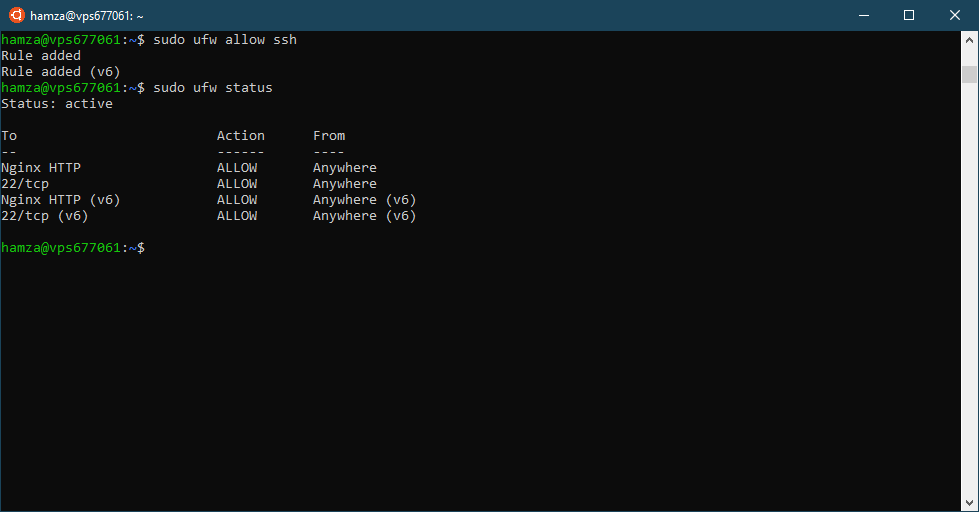


Figure 18 - Autorisation du ssh à travers le pare-feu

* Vérification de l’état de nginx

La commande suivante nous indique l’état du nginx.

$ systemctl status nginx

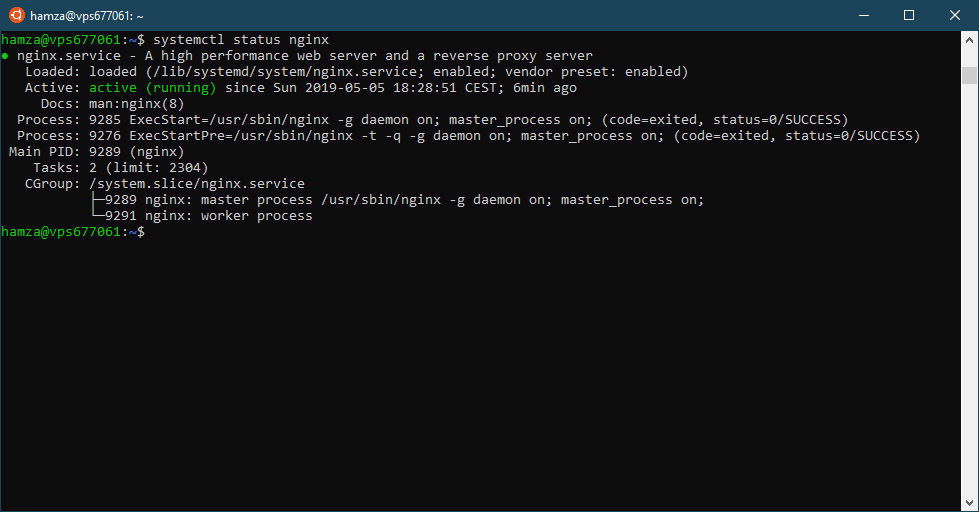


Figure 19 - État du service nginx

* Configuration de nginx

La capture d’écran suivante indique la configuration nécessaire pour nginx pour gérer la communication entre le monde extérieur (les requêtes venant de l’extérieur) et le service de détection des pathologies. En effet c’est un fichier qui inclut les règles qui spécifie à nginx comment agir.

$ vim /etc/nginx/sites-available/padvpas-rest.conf

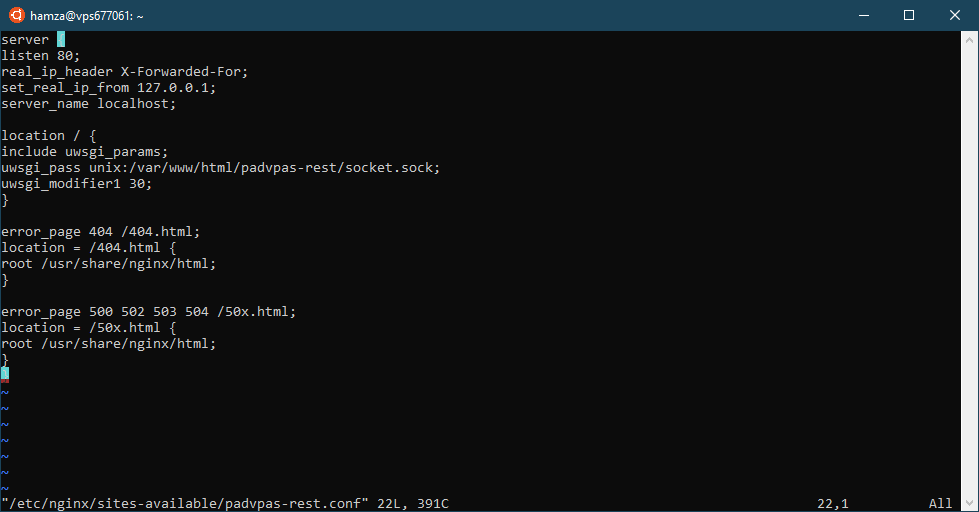


Figure 20 - Fichier de configuration de nginx

Ensuite une bonne pratique exige la création d’un lien symbolique vers le fichier de configuration dans le dossier /etc/nginx/sites-enbled/ en effet, c’est à partir de ce répertoire nginx charge le fichier de configuration.

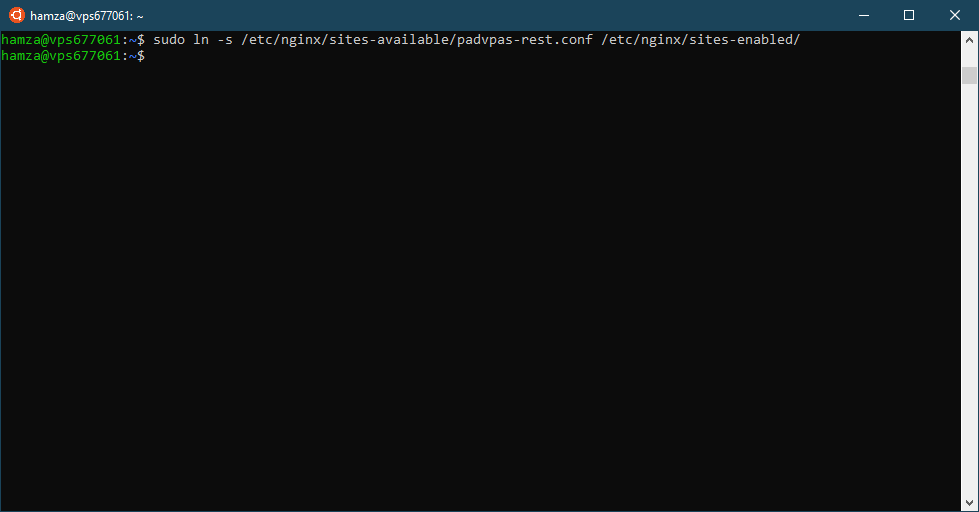


Figure 21 - Création d'un lien symbolique vers le fichier de configuration

* Créer un répertoire pour contenir le code du service de détection des pathologies

$ mkdir /var/www/html/padvpas-rest

* Clonage du code du service de détection

On à téléchargé le code du service de détection des pathologies vocales vers GitHub[[3]](#footnote-3), pour qu’on puisse le copier à notre VPS facilement…

$ git clone <https://github.com/Hamza-Hedfi/PADVPAS_API.git>

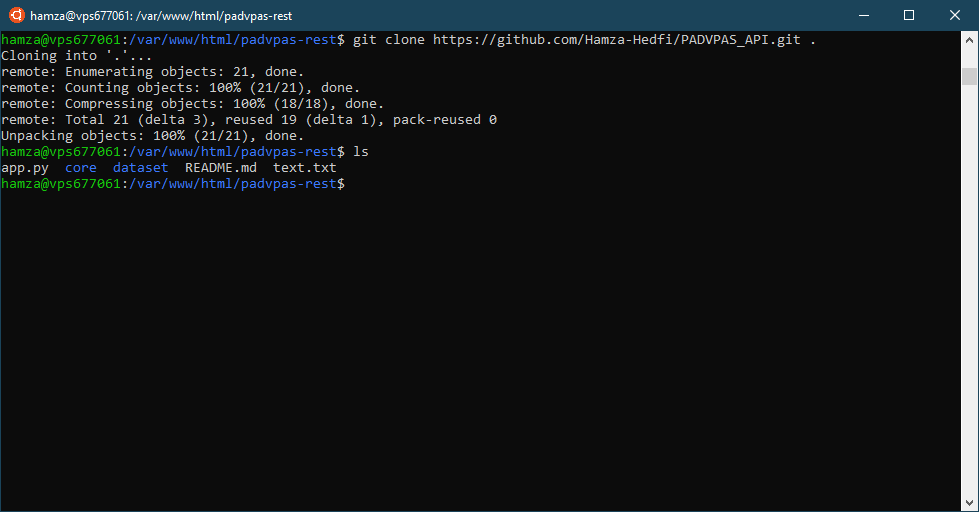


Figure 22 - Clonage du code

* Installation de Python

A partir de cette étape on va s’intéresser au service de détection des pathologies vocales (installation des packages nécessaires à son fonctionnement, configuration, …)

On a besoin de python et d’autres packages (build-essential, …), la commande suivante permet l’installation de ces derniers.

$ sudo apt install build-essential python3.7 python3.7-dev

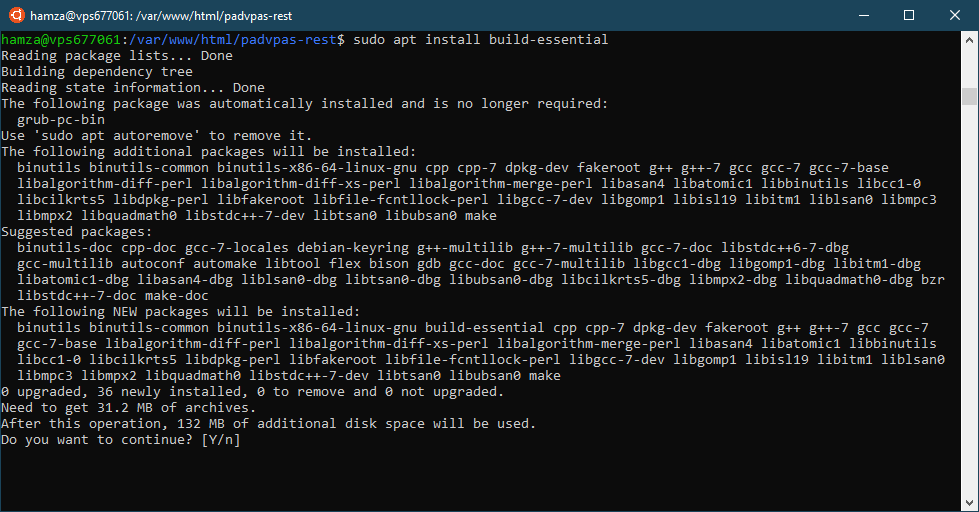


Figure 23 - Installation du package build-essential

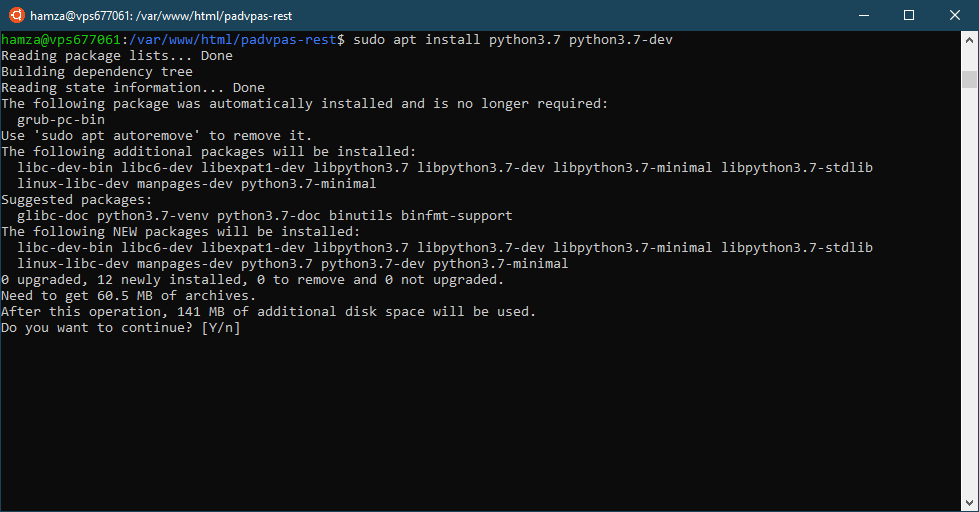


Figure 24 - Installation du package python3.7 et python3.7-dev

* Création d’un environnement virtuel

Un environnement virtuel est un outil permettant de séparer les dépendances requises par différents projets en leur créant des environnements virtuels python isolés.

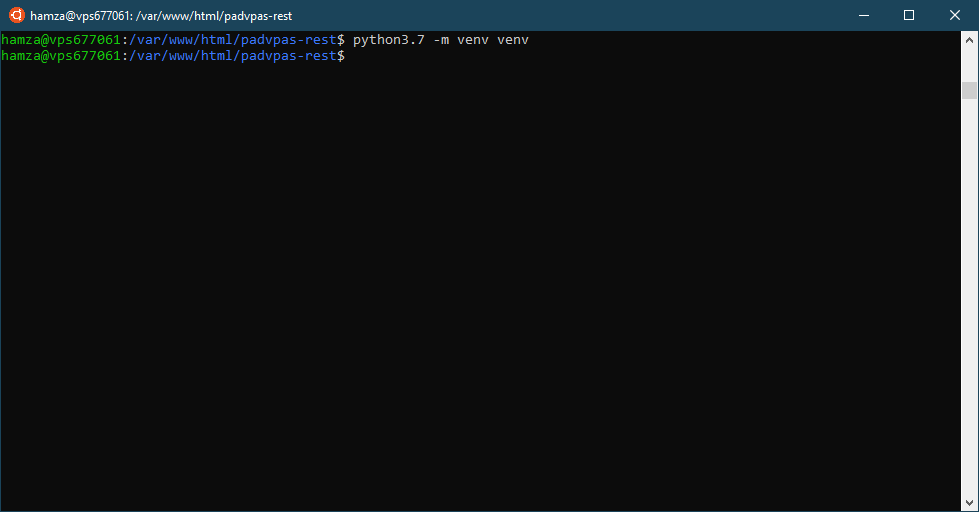


Figure 25 - Création d'un environnement virtuel

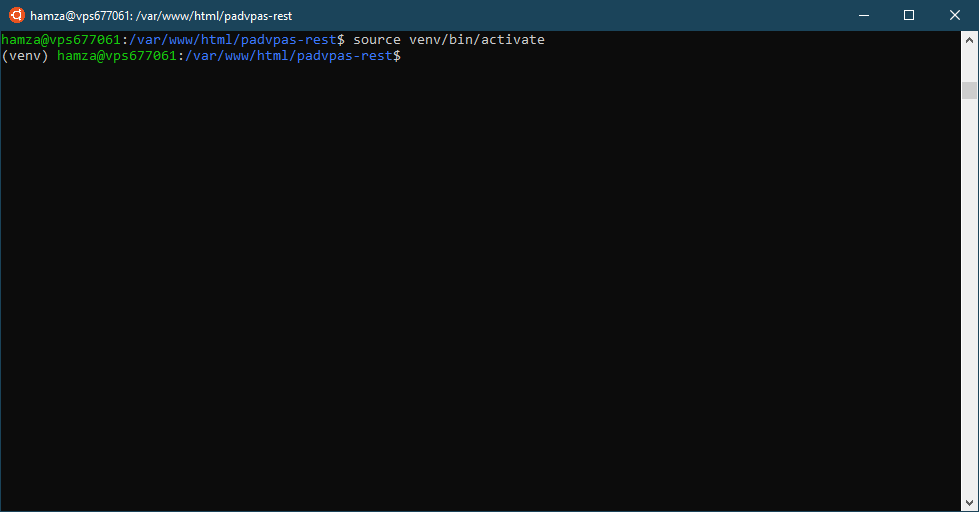


Figure 26 - Activation de l'environnement virtuel

* Installation des packages nécessaire au fonctionnement du service de détection des pathologies vocales

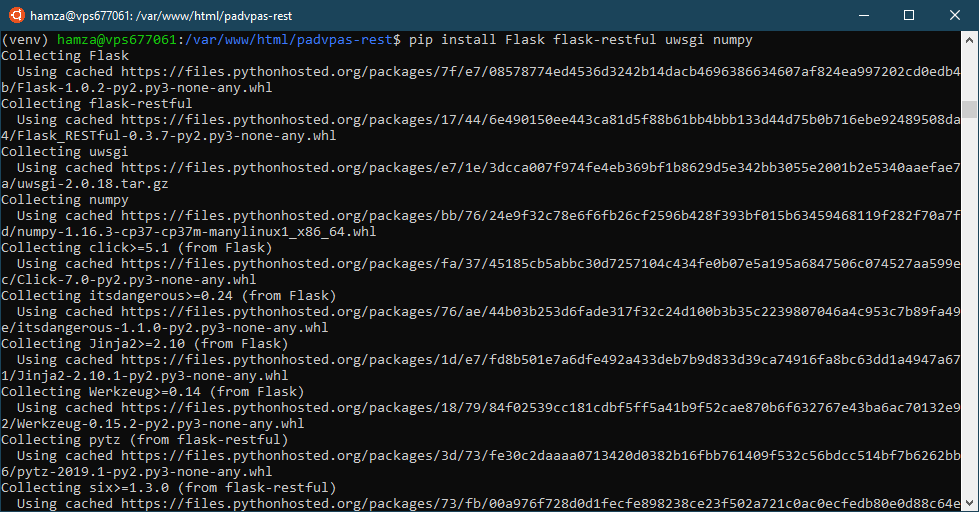


Figure 27 - Installation de Flask, FlaskRESTful, uwsgi, numpy

* Création du service

La capture d’écran suivante indique la configuration pour le service de détection des pathologies vocales.

$ sudo vim /etc/systemd/system/uwsgi\_padvpas\_rest.service

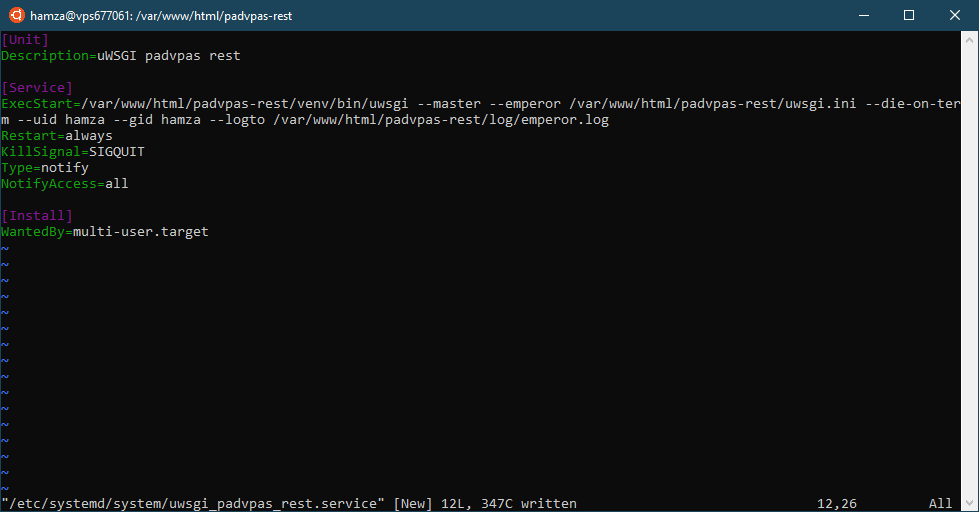


Figure 28 - Fichier de configuration de service

Ensuite on indique la configuration de uWSGI[[4]](#footnote-4)

$ vim /var/www/html/padvpas-rest/uwsgi.ini

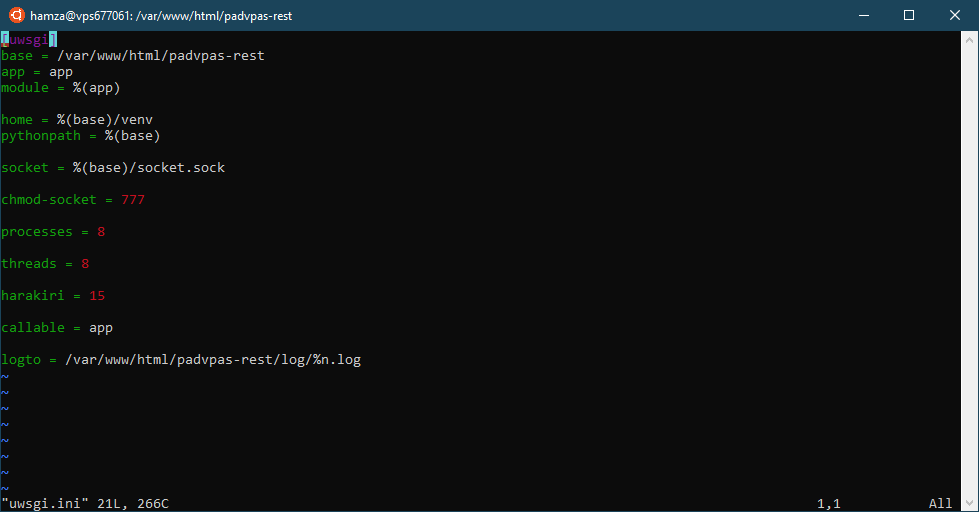


Figure 29 - Fichier de configuration de uwsgi

* Démarrage du service

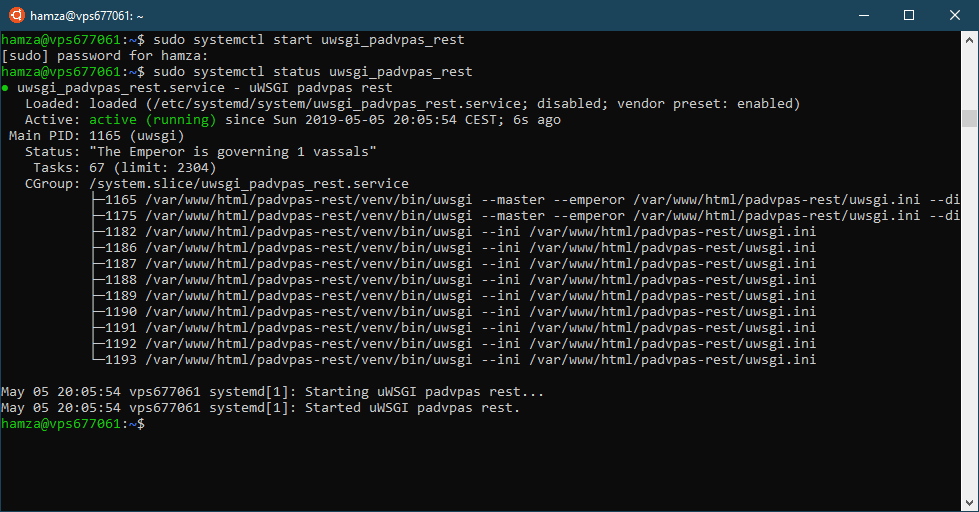


Figure 30 - Démarrage et vérification de l'état du service

* Test

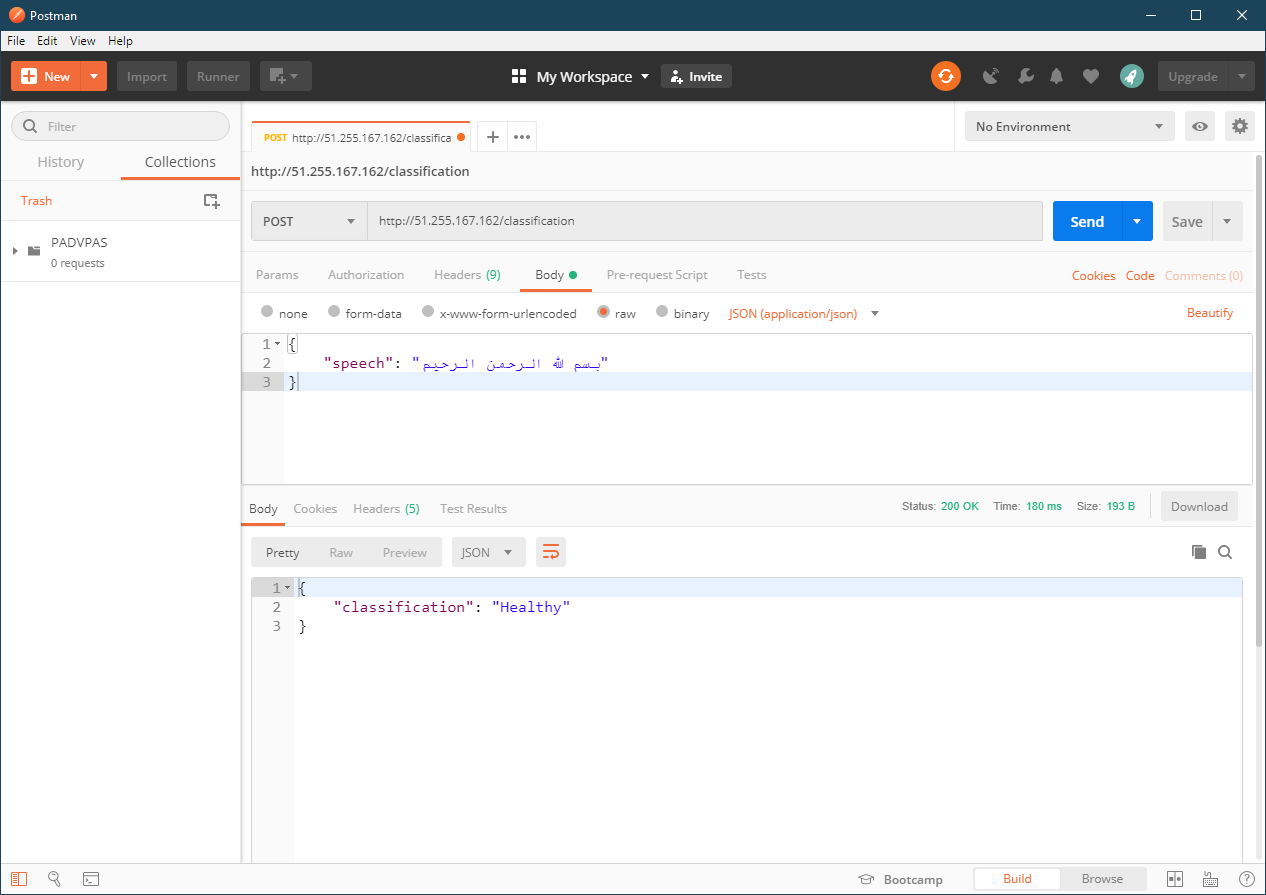


Figure 31 - Première test du service de détection des pathologies

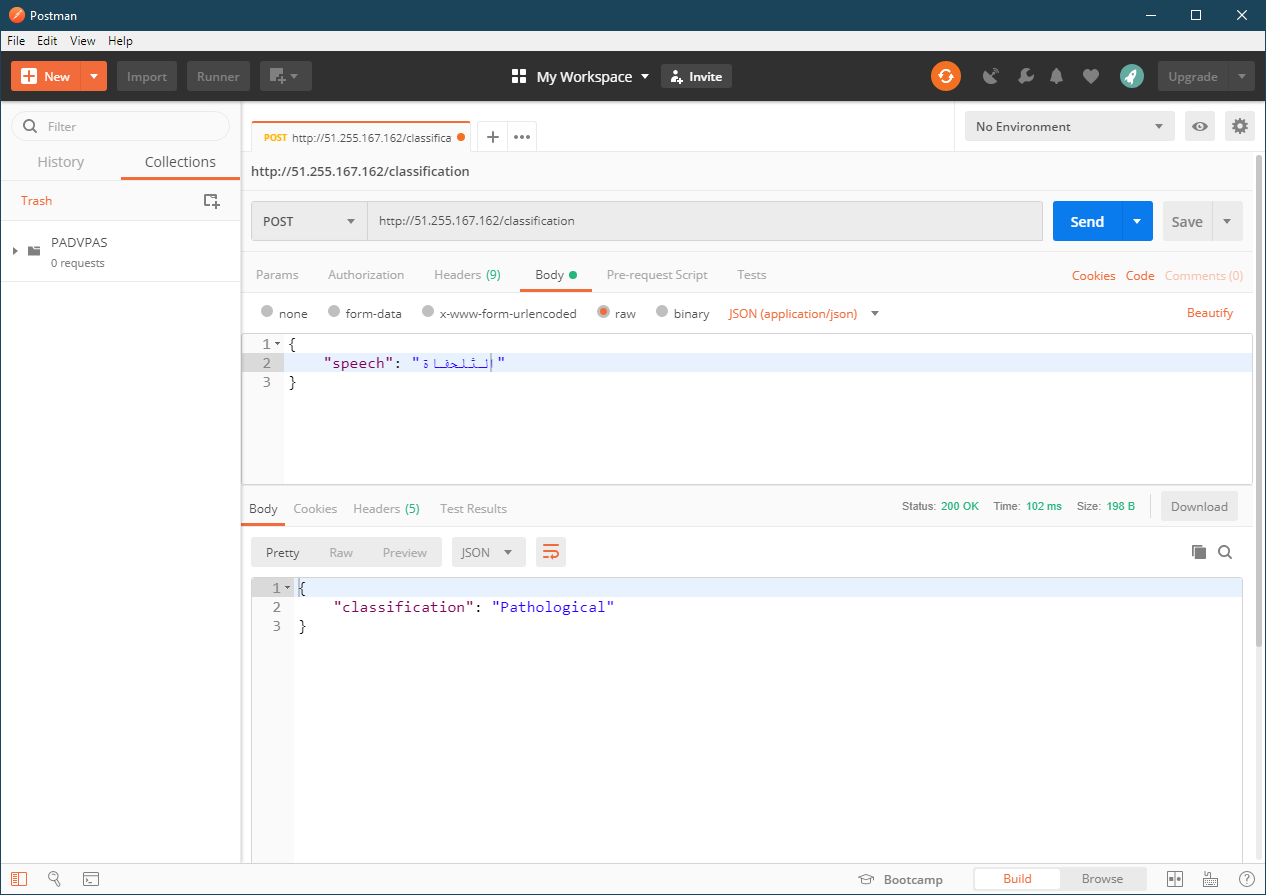


Figure 32 - Deuxième test du service de détection des pathologies

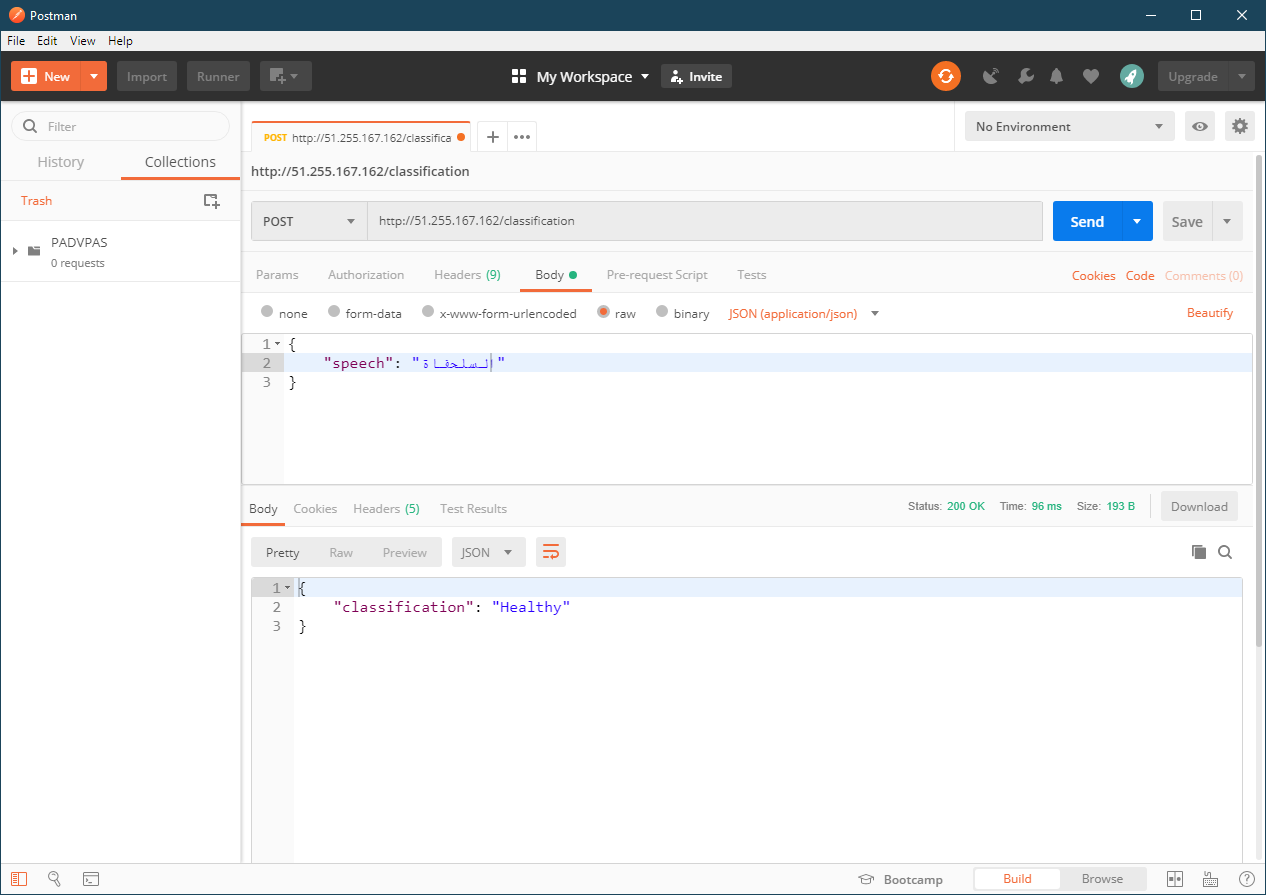


Figure 33 - Troisième test du service de détection des pathologies

## 4. Le modèle acoustique

Un modèle acoustique est utilisé dans la reconnaissance automatique de la parole pour représenter la relation entre un signal audio et les phonèmes ou autres unités linguistiques qui représentent la parole. Le modèle est appris à partir d'un ensemble d'enregistrements audio et de leurs transcriptions correspondantes. Il est créé en prenant des enregistrements audios de la parole et leurs transcriptions de texte, pour créer des représentations statistiques des sons qui composent chaque mot.

### 4.1 Pourquoi entrainer un modèle acoustique ?

Ce modèle acoustique est vital pour le bon fonctionnement de notre application. Mais comment vital ?

La méthode pour la détection des pathologies vocales dans le discours arabe, se base sur la transcription de la parole, c’est à travers la version textuelle de la parole qu’on fait les calculs nécessaires à la détection des pathologies.

Sans la représentation identique de la parole sous sa forme écrite, on ne peut pas détecter les pathologies contenues dans la parole produite.

Le module ou bien package de la reconnaissance automatique de la parole utilisé est Pocketshpinx. La communité de CMUsphinx offre une panoplie des modèles acoustiques pré-entrainés pour des diverses langues, mais l’arabe n’était pas inscrite dans cette panoplie. La capture d’écran suivante résume la liste des modèles acoustiques disponibles.

S’il y avait un modèle acoustique pour l’arabe, on aurait pu l’adapter selon notre besoin.

Dans tous les cas l’entrainement d’un modèle acoustique est obligatoire.

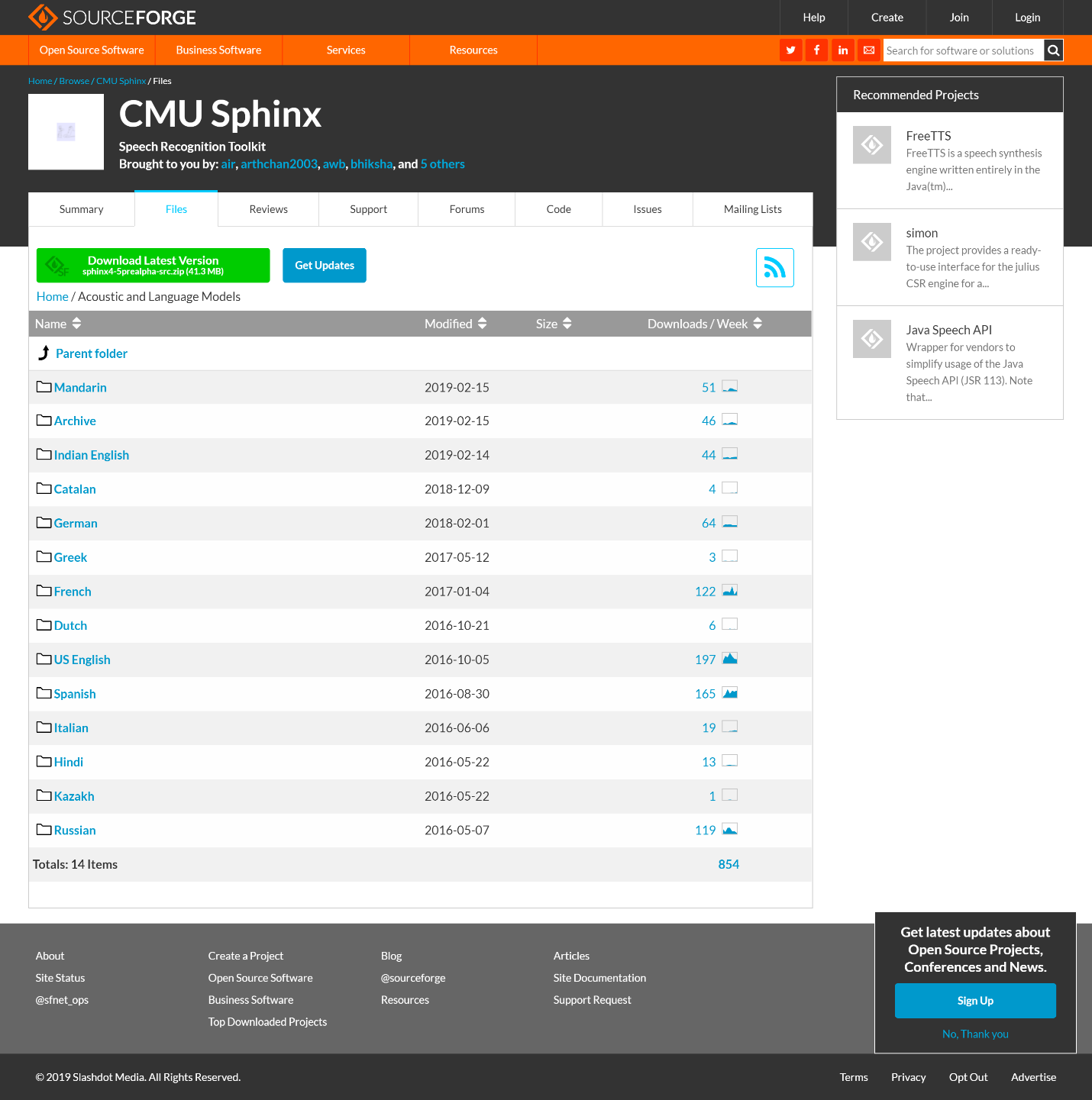


Figure 34 - la liste des modèles acoustiques disponibles (https://sourceforge.net/projects/cmusphinx/files/Acoustic%20and%20Language%20Models/ )

### 4.2 Entrainement du modèle acoustique

Pour une description complète de la phase d’entrainement du modèle acoustique s’il vous plaît se référer à la documentation de CMU Shpinx (<https://cmusphinx.github.io/wiki/tutorialam/>)

## 5. L’application mobile

Dans cette section je vais détailler brièvement l’application mobile.

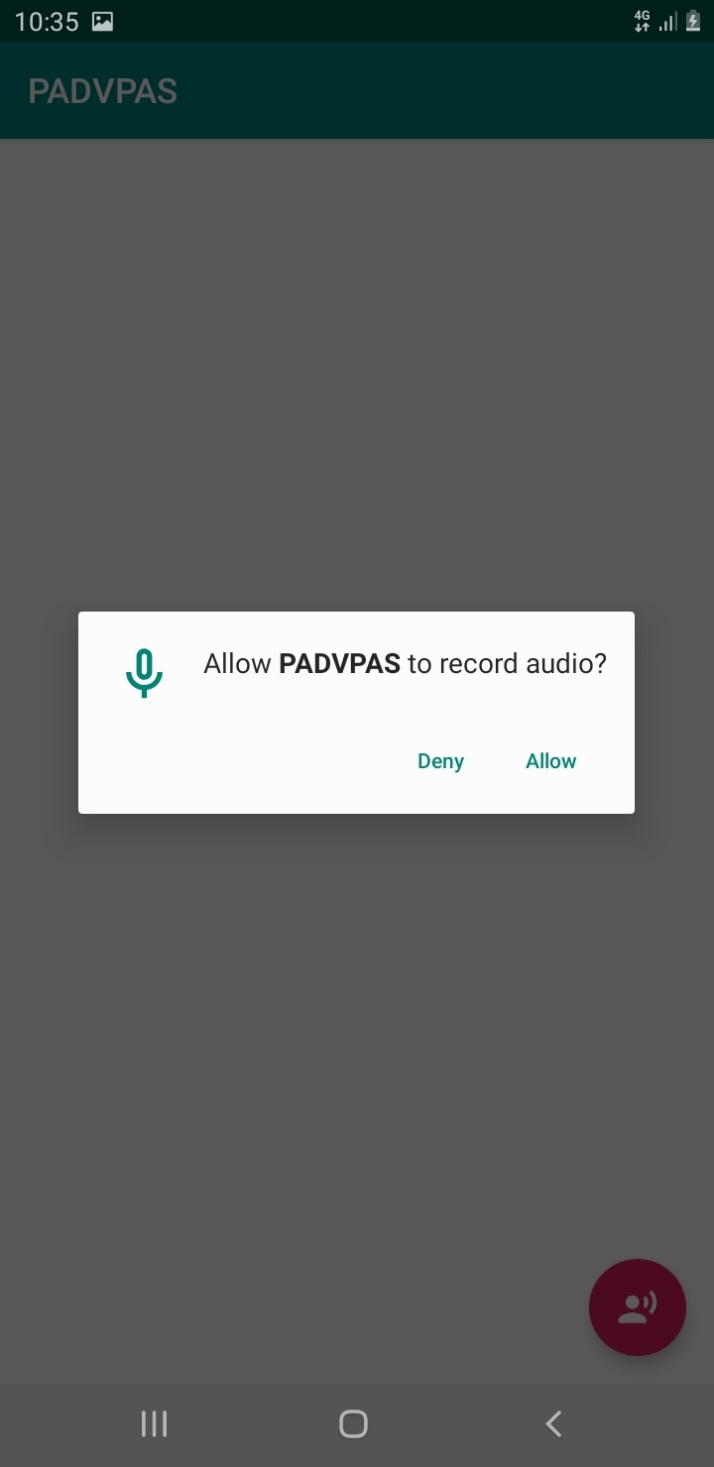
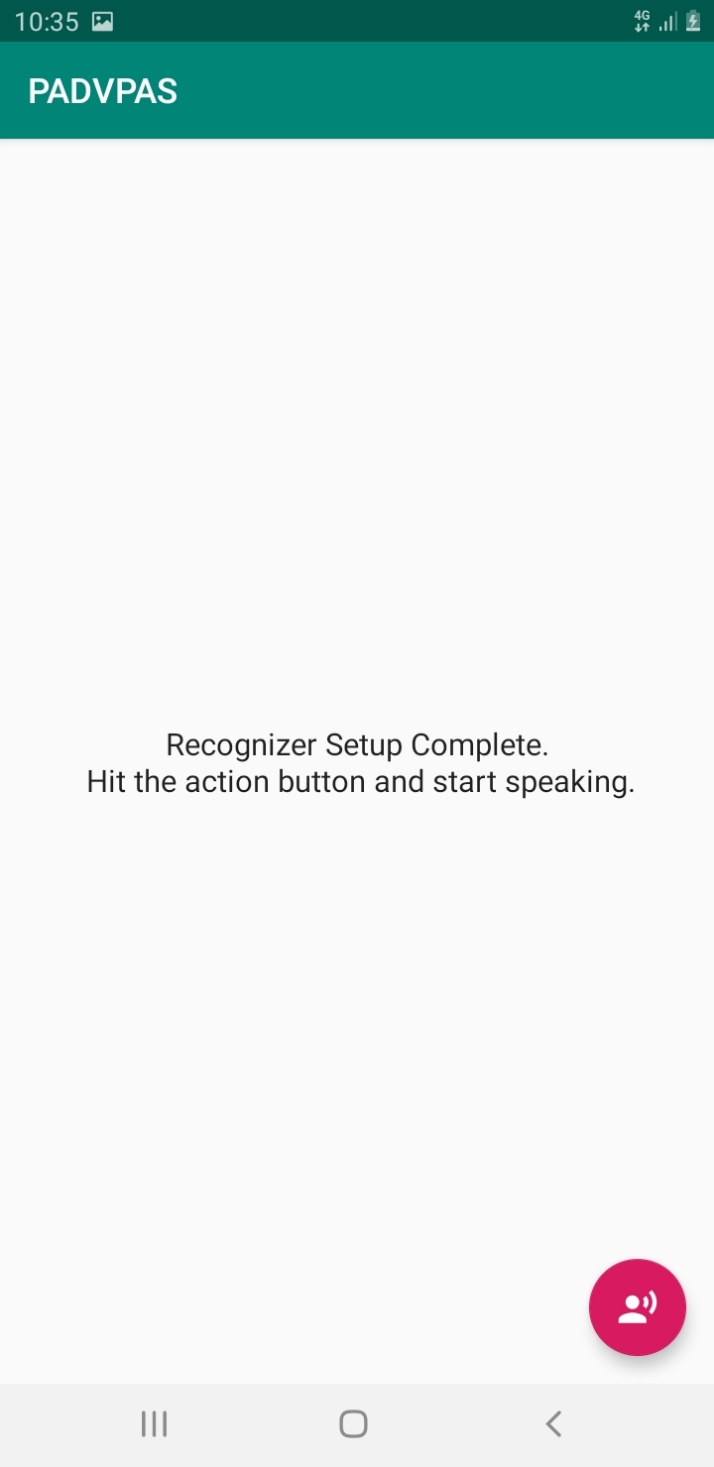
Tout d’abord la reconnaissance automatique de la parole se fait au niveau de l’application mobile à l’aide de Pocketsphinx. Une phase d’initialisation du Pocketsphinx est déclenchée à la première exécution de l’application. Ensuite après l’autorisation d’utilisation de microphone de l’appareil l’application est prête pour l’utilisation.

Figure 35 - Première exécution de l'application

Figure 36 - Application prête à être utiliser

L’utilisateur commence à parler, la transcription de sa parole est affichée, et d’ici la communication avec le service de détection des pathologies vocales commence.

Figure 37 - Application en écoute

Figure 38 - Aucun parole détecté

Après avoir reçu les résultats on mit à jour l’interface pour indique si la parole est pathologique ou non.

Figure 39 – Résultat obtenu (cas de la parole pathologique)

Figure 40 -Résultat obtenu (cas de la parole saine)

Conclusion et Perspectives

Conclusion

L’élaboration de notre application nous a permis de nous confronter à une expérience professionnelle dans le domaine de développement des applications que nous ne pouvons qualifier qu’enrichissante. En outre, elle nous a offert l’opportunité de concrétiser nos connaissances théoriques à l’aide d’un cas réel. Ainsi, ce projet nous a apporté plus concernant les langages de programmations et les plateformes que nous avons découverts et pratiqués.

Dans ce travail, nous avons abordé le problème de la détection des pathologies vocales contenues dans le discours arabe. Pour ce faire, nous avons présenté, en premier lieu, la reconnaissance automatique de la parole, les différents ordres de difficultés de la compréhension de la parole humaine lors de l’interaction homme/machine. Nous avons aussi soulevé les différents facteurs qui peuvent engendrer des pathologies vocales.

À titre général, ce stage a constitué une expérience professionnelle exceptionnelle. La vie universitaire ne permet que d’avoir un aperçu de son futur métier. On découvre rapidement que l’aspect pratique dépasse souvent les connaissances acquises et qu’il constitue une formation rapide.

Perspectives

Comme tout travail, ce projet de fin d’études présente quelques limites qui nécessitent des études plus approfondies pour les dépasser. En effet, nous souhaitons, agrandir la taille des ressources utilisées (principalement le dictionnaire de prononciation) pour améliorer le taux de détection.

Nous souhaitons également étendre l’application pour identifier le type de la pathologie vocale (phonèmes problématiques, facteur origine de la dégradation de la parole), et la corriger.

A long terme, nous souhaitons réaliser une application d’apprentissage de la langue arabe.

# Glossaire

**Arpa** : (Acronyme de Advanced Research Projects Agency, Agence pour les projets de recherche avancés ). Organisme des Etats-unis remplacé par le DARPA qui a soutenu énormément de projets informatiques historiques.

**Bit** : Abréviation de «BInary digiT», mais bit signifie «un peu», «un bout» en anglais. Unité élémentaire d'information, ne pouvant prendre que deux valeurs, représentées par 0 et 1 en général (ou Faux et Vrai).

**Echantillonnage** : Transformer un signal analogique, capturé par intervalle de temps, en signaux numériques équivalents du point de vue de l'information transmise, pour qu'ils soient manipulés par ordinateur.

**DTMF**: (Acronyme de Dual Tone Multi Frequency, Signalisation multifréquence à double tonalité). Le système de signalisation à fréquence vocale est utilisé pour transmettre des données de signalisation par téléphone ou par radio. Un téléphone à touches génère les tonalités DTMF pendant la numérotation.

**DTW** : (Acronyme de Dynamic Time Warping, déformation dynamique temporelle). Algorithme qui permet de trouver le meilleur appariement entre une référence enregistrée et un signal à reconnaître, par le calcul d'une différence entre des vecteurs respectifs de caractéristiques.

**MMC, Modèles de Markov cachés** : (HMM en anglais). Du nom du mathématicien russe. L’introduction de cette théorie probabiliste dans le champ de la reconnaissance vocale a permis de représenter des sons élémentaires sous forme statistique.

**Numériser** : Transformer un signal quelqu’il soit (analogique, optique, etc…) en un signal numérique pour qu'il soit manipulé par un ordinateur.

**Phonème** : Unité qui permet de caractériser tous les sons d’une langue. La plupart des langues comportent moins d’une centaine de phonèmes. Par exemple «a», «è», «ss», etc…

**RAP** : Reconnaissance automatique de la Parole(ou ASR : Automatic Speech Recognition, en anglais).

**SAP** : Synthèse Automatique de la Parole (ou TTS : Text To Speech, en anglais)

**Spectrogramme** : Le spectrogramme est une représentation qui permet d'analyser des portions successives d'un signal. Pour cela le signal est considéré comme stationnaire à court-terme sur une fenêtre d'analyse définissant la portion analysée de façon à ce que l'analyse spectrale classique soit localement valable.

# Liste bibliographique

[1] Méloni H. Fondements et perspectives en traitement automatique de la parole, Ed. AUPELF-UREF, Décembre 1996

[2] Terbeh, N., Maraoui, M., Zrigui, M.: Probabilistic Approach for Detection of Vocal Pathologies in the Arabic Speech. CICLing 2015, April 14-20, 2015, Cairo-Egypt, 2015.

[3] Remacle A. La charge vocale : De sa quantification à l’étude de son impact sur la fonction phonatoire et sur la qualité vocale, Thèse de doctorat, Université de Liège

[4] Hosom, J.-P., Kain, A. B., Mishra, T., Van Santen, J. P., Fried-Oken, M. andStaehely, J. Intelligibility of modifications to dysarthric speech. In Proceedings of theIEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP), 878–881.

[5] Howard, D. M., and Angus, J. A. S. Room acoustics: How they affect vocal production and perception. In P. H. Dejonckere (Ed.), Occupational voice: Care and cure,29-46

1. Nginx est un serveur Web qui peut également être utilisé en tant que proxy inverse, équilibreur de charge, proxy de messagerie et cache HTTP. [↑](#footnote-ref-1)
2. Uncomplicated Firewall (UFW) est un programme de gestion de pare-feu Netfilter conçu pour être facile à utiliser. Il utilise une interface de ligne de commande composée d'un petit nombre de commandes simples et utilise iptables pour la configuration. UFW est disponible par défaut dans toutes les installations Ubuntu après 8.04 LTS. [↑](#footnote-ref-2)
3. GitHub est un service web d'hébergement et de gestion de développement de logiciels, utilisant le logiciel de gestion de versions Git. [↑](#footnote-ref-3)
4. Web Server Gateway Interface est une convention d'appel simple permettant aux serveurs Web de transférer les demandes aux applications Web ou aux infrastructures écrites en langage de programmation Python. [↑](#footnote-ref-4)