



SPEECHAPP : un logiciel de reconnaissance automatique de la parole

[projet du MIG *Systèmes Embarqués* 2013]

Julien CAILLARD, Adrien DE LA VAISSIÈRE, Thomas DEBARRE,
Matthieu DENOUX, Maxime ERNOULT, Axel GOERING,
Clément JOUDET, Nathanaël KASRIEL, Anis KHLIF,
Sofiane MAHIOU, Paul MUSTIÈRE, Clément ROIG, David VITOUX

18/11/13 - 6/12/13



Remerciements

Toute l'équipe SE souhaiterait tout d'abord remercier Valérie Roy pour son engagement, ses bons conseils et son dynamisme. +remercier tout les intervenants dont je n'ai pas la liste... (Catherine, le jury, IRIS, Eurocopter avec Khodor, Dassault, Esterel, Coq, la bibliothécaire)

Table des matières

0.1	Présentation des enjeux	5
0.2	Objectifs du projet	5
0.3	Approches de la reconnaissance vocale	6
0.3.1	Acoustique-phonétique	6
0.3.2	Reconnaissance de motifs	6
I	Démarche Technique	8
1	Principe général du traitement du signal	8
1.1	Objectifs	8
1.2	Schéma global	8
2	Analyse, formatage du signal	10
2.1	Introduction	10
2.2	Prérequis	10
2.2.1	Qu'est-ce que le son ?	10
2.2.2	Comment le son est-il représenté dans l'ordinateur ?	10
2.3	Enregistrement, recadrage, filtrage HF	12
2.3.1	Synchronisation	12
2.3.2	Filtrage passe-haut	12
2.4	Echantillonnage, fenêtrage	14
2.5	Transformée de Fourier	15
2.6	Simulation du comportement de l'oreille humaine	16
2.7	Transformée inverse ou Décomposition en cosinus inverse	17
3	Modélisation des mots à reconnaître par les modèles de Markov cachés	19
3.1	Objectifs	19
3.2	Prérequis et principe	19
3.2.1	Les automates	19
3.2.2	Les modèles de Markov cachés	19
3.2.3	Application à la reconnaissance vocale	20
3.3	Principaux algorithmes sur les modèles de Markov	20
3.4	Application à notre objectif	21
3.5	Phase d'apprentissage	21
3.6	Phase de reconnaissance	22
II	Approche commerciale	24
1	Approche du développement du projet	24
1.0.1	Choix d'une architecture optimale pour notre projet	24
1.0.2	Réalisation du SpeechServer	26
1.0.3	Système de Gestion de Base de Données (SGBD)	27
1.1	Dimensionnement de l'infrastructure de calcul de The Speech App Company	27

1.1.1	Hypothèses de fonctionnement	27
1.1.2	Dimensionnement en mémoire RAM et espace disque	28
1.1.3	Dimensionnement réseau	28
1.1.4	Dimensionnement des éléments de calculs	28
1.1.5	Choix de l'infrastructure et coûts liés	28
1.1.6	SpeechRecorder	29
1.1.7	SpeechApp	30
2	Applications	33
3	Budget, modèle économique	35
3.1	Introduction	35
3.2	Les salaires	35
3.3	Le compte de résultat prévisionnel	36
3.4	Le bilan	37
3.5	Les impôts	37
3.6	Conclusion et vue sur le long terme	38
III	Code	40
A	Code Principal	40
A.1	shell.py	40
A.2	server.py	46
A.3	gui.py	47
B	handling	53
B.1	fenetrehann.py	53
B.2	inverseDCT.py	54
B.3	triangularFilterbank.py	54
B.4	passehaut.py	55
B.5	fft.cpp	55
C	HMM	58
C.1	creationVecteurHMM.py	58
C.2	markov.py	59
C.3	tableauEnergyPerFrame.py	63
C.4	hmm.cpp	64
D	recorder	83
D.1	recorder.py	83
D.2	sync.py	84
E	utils	87
E.1	animate.py	87
E.2	constantes.py	87
E.3	db.py	88
E.4	util.py	93
F	SpeechApp	94
F.1	main	94
F.2	holder	99
F.3	recorder	110
F.4	recorderWorker	111
F.5	index.html	114

G	SpeechServer	117
G.1	main.py	117
G.2	audioConverter.py	119
G.3	clientAuth.py	121
G.4	speechActions.py	123

Introduction

0.1 Présentation des enjeux

La reconnaissance vocale automatisée est l’objet d’intenses recherches depuis plus de 50 ans. Malgré son caractère d’abord futuriste, comme cela peut se retrouver dans de nombreuses oeuvres de science-fiction, elle a pris sa place dans nos quotidiens avec la prolifération de systèmes qui embarquent une telle technologie, par exemple avec le logiciel Siri dans les téléphones d’Apple[1]. Les perspectives économiques qui s’ouvrent au détenteur d’un système de reconnaissance fiable, robuste, et portable sont innombrables et l’on ne saurait surestimer son importance, (systèmes embarqués, commandes vocales, aide aux sourds/muets, ...). Les derniers systèmes les plus aboutis offrent des performances remarquables, mais le problème reste toujours ouvert et suscite plus d’engouement que jamais en raison de la croissante puissance de calcul disponible et les dernières avancées et applications découvertes.

La complexité de ce problème s’explique notamment par la grande diversité des thèmes qui lui sont connexes et que tout système se voulant performant se doit d’incorporer (traitement du signal, théorie de l’information, acoustique, linguistique, intelligence artificielle, physiologie, psychologie, ...). La reconnaissance vocale requiert des connaissances trop diverses pour être maîtrisées par un seul individu et la capacité à savoir exploiter des ressources dont on est pas expert devient un atout capital. Elle ne se réduit pas à la seule détermination d’une suite de mots prononcés, mais peut s’étendre à divers autres applications telles que la reconnaissance de langage, d’accent, déterminer le sexe et l’âge du locuteur, s’il est stressé ou calme, dans quel environnement il se trouve, tant ces paramètres influent de manière capitale sur l’analyse.

0.2 Objectifs du projet

Ce MIG s’est placé dans une perspective résolument plus humble en raison du temps imparti. Il ne s’agissait pas de réaliser un programme prétendant rivaliser avec les actuels systèmes de reconnaissance, fruits de nombreuses années de recherches et de développement ; mais plutôt, à l’instar de l’ingénieur généraliste, de prendre connaissance d’un sujet et d’une problématique et tâcher, en équipe, d’y apporter une solution qui soit la plus optimale possible compte tenu des exigences temporelles et matérielles. Le projet des MIG ne se réduisant pas non plus à une réalisation technique il s’agissait de garder en vue les perspectives économiques et les composantes juridiques, indissociables d’un tel projet, comme garde fou de toute pérégrination informatique.

De plus, ce système de reconnaissance vocale, qui peut sembler immédiat tel qu’on l’expérimente aujourd’hui, n’est en fait pas si évident qu’il y paraît. En témoigne la faible réussite de ces applications en général puisque nous avons tous ressenti un jour la frustration de ne pas être compris de la machine. Il convient donc de préciser ce qui rend la tâche si subtile face à ce que nos oreilles et notre cerveau fait aussi instantanément.

Les rôles ont été attribués dès le début selon les goûts et compétences de chacun mais la pertinente répartition des tâches, la diversité intrinsèque au projet et l’angle avec lequel nous l’avons abordé a permis à chacun d’exploiter un panel très diversifié de ses compétences tout en apportant la valeur ajoutée de sa spécialité. Chaque fonction dépendant très fortement de ce qui précède et de ce qui

suit, une bonne communication interne était indispensable pour un développement juste et efficace. Si la coordination spontanée d’une équipe de treize personnes a été au début délicate, une indéniable rigueur et discipline adjointe à l’exploitation de ressources adaptées ont vite imposé une organisation naturelle. Par exemple l’utilisation de la plateforme github[2] pour l’échange de fichiers et de mises à jour s’est révélée particulièrement efficace et permettait à chacun d’incorporer en temps réels les dernières modifications. La complexité de la discipline fut un des principaux obstacles, et une phase d’appropriation des techniques requises, de part la lecture de livres dédiés, d’articles de recherches ainsi que de thèses a été le poumon du projet. Le caractère abscon de certains articles a rajouté à la difficulté.

0.3 Approches de la reconnaissance vocale

Avant de rentrer dans des considérations techniques, il est nécessaire de définir un principe d’étude, une stratégie de résolution qui dictera l’orientation générale du projet en plus de rendre les objectifs et les enjeux plus clairs. Cette partie a pour but de donner un aperçu des différents angles d’attaques du problème donné pouvant être considérés, ainsi que de présenter celui que nous avons choisi, avec quelles motivations.

Dans son livre *Fundamentals of speech recognition*, Lawrence Rabiner[3] dégage des travaux de ces prédécesseurs trois approches conceptuelles du problème. Ces approches sont les suivantes : l’approche acoustique-phonétique, l’approche par reconnaissance de motifs et l’approche par intelligence artificielle. Cette dernière n’étant, d’après Rabiner, qu’un avatar de la première ; nous ne présenterons que l’acoustique phonétique et la reconnaissance de motifs que nous avons choisi pour notre projet.

0.3.1 Acoustique-phonétique

L’approche acoustique-phonétique est indubitablement celle qui paraît la plus naturelle et directe pour faire de la reconnaissance vocale et est celle qui s’impose a priori à l’esprit. Le principe est le suivant : l’ordinateur tâche de découper l’échantillon sonore de manière séquentielle en se basant sur les caractéristiques acoustiques observées et sur les relations connues entre caractéristiques acoustiques et phonèmes. Ceci dans le but d’identifier une suite de phonèmes¹

Cette approche suppose qu’il existe un ensemble fini de phonèmes différentiables et que leurs propriétés sont suffisamment manifestes pour être extraites d’un signal ou de la donnée de son spectre (tableau des fréquences et de leur amplitude associée, composant un signal à un instant donné) au cours du temps. Même si il est évident que ces caractéristiques dépendent très largement du sujet parlant, on part du principe que les règles régissant la modification des paramètres peuvent être apprises et appliquées.

Bien qu’elle ait été vastement étudiée et soit viable on lui préférera l’approche par reconnaissance de motifs qui, pour plusieurs raisons, l’a supplantée dans les systèmes appliqués. C’est celle que nous avons choisi et que nous présentons dans le prochain paragraphe.

0.3.2 Reconnaissance de motifs

Cette technique diffère de la méthode précédente par le fait qu’elle ne cherche pas à exhiber des caractéristiques explicites. Elle se compose de deux étapes : « l’entraînement » des motifs, et la reconnaissance via la comparaison de ces motifs.

1. « En phonologie, domaine de la linguistique, un phonème est la plus petite unité discrète ou distinctive (c’est-à-dire permettant de distinguer des mots les uns des autres) que l’on puisse isoler par segmentation dans la chaîne parlée. Un phonème est en réalité une entité abstraite, qui peut correspondre à plusieurs sons. Il est en effet susceptible d’être prononcé de façon différente selon les locuteurs ou selon sa position et son environnement au sein du mot. et d’ainsi reconnaître un mot. » (Définition Wikipédia du mot phonème)

L'idée sous-jacente au concept d'entraînement repose sur le principe selon lequel si l'on dispose d'un ensemble suffisamment grand de version d'un motif à reconnaître, on doit être capable de caractériser pertinemment les propriétés acoustiques du motif. Notons que les motifs en question peuvent être de nature très diverses, comme des sons, des mots, des phrases ; ce qui sous-tend l'idée d'un grand nombre d'applications théoriques comme présenté en introduction. La machine apprend alors quelles propriétés acoustiques sont fiables et pertinentes. On effectue ensuite une comparaison entre le signal à reconnaître et les motifs tampons, afin de le classifier en fonction du degré de concordance.

Sans plus entrer dans les détails, les avantages de cette approche qui nous ont poussés à l'adopter sont les suivants :

- Elle est simple à appréhender, et est très largement comprise et utilisée
- Elle est robuste, c'est-à-dire qu'elle dépend peu du locuteur et de l'environnement
- Elle donne lieu à de très bons résultats

Première partie

Démarche Technique

1. Principe général du traitement du signal

1.1 Objectifs

Bien que la reconnaissance vocale telle qu'elle est aujourd'hui mise-en-place dans les différents matériels semble immédiate, le travail à effectuer pour reconnaître un mot est complexe. La première étape pour faire de la reconnaissance vocale est de parvenir à trouver un moyen de caractériser efficacement et uniformément un mot. Cela désigne un mot par un certain motif puis permet par le même procédé appliqué sur un enregistrement quelconque, de parvenir à identifier deux motifs proches qui correspondraient alors au même mot. Il s'agit donc tout d'abord de traiter le signal pour en découvrir certaines caractéristiques. En effet, une même personne ne prononce pas toujours les mots de la même façon, au même débit, avec les mêmes hauteurs de son, ce qui rend ardue une simple identification par comparaisons temporelles.

1.2 Schéma global

Afin de gérer ces difficultés, nous avons mis en place plusieurs étapes de traitement supplémentaires afin d'obtenir cette fameuse « trace » qui caractériserait un enregistrement, c'est-à-dire un mot. Nous avons pour cela utilisé plusieurs techniques de traitement du signal communément connues (échantillonnage, fenêtrage, transformée de Fourier directe et inverse). Cette figure explique globalement le traitement que nous avons choisi de mettre-en-place afin de reconnaître le mot prononcé. Il y a donc plusieurs étapes qui s'enchaînent pour parvenir à un objet que nous pourrions manipuler en le sachant représentatif et caractéristique du son.

Enregistrement du son La première étape consiste simplement à enregistrer le son sur le disque dur de l'ordinateur. Nous utilisons pour cela un module intégré à Python appelé PyAudio[4]. Cela permet d'enregistrer avec une certaine fréquence d'échantillonnage (donc un certain nombre de captures de son par seconde) les amplitudes du son captées par le micro.

Découpage en fenêtre Le son est découpé ensuite en petites fenêtres de quelques dizaines de millisecondes ce qui permet d'isoler les événements sonores qui pourraient avoir une importance. Il s'agit d'un *fenêtrage*.

Passage en fréquence Jusque là, le son étudié se représentait temporellement ce qui avait été entendu. Néanmoins, il est difficile d'étudier un son tel quel et on utilise alors le lien entre les fréquences et le signal temporel. Il est ensuite plus facile d'étudier et de transformer un ensemble de fréquences pour appliquer par un exemple des filtres qui rapprochent le programme du fonctionnement de l'oreille.

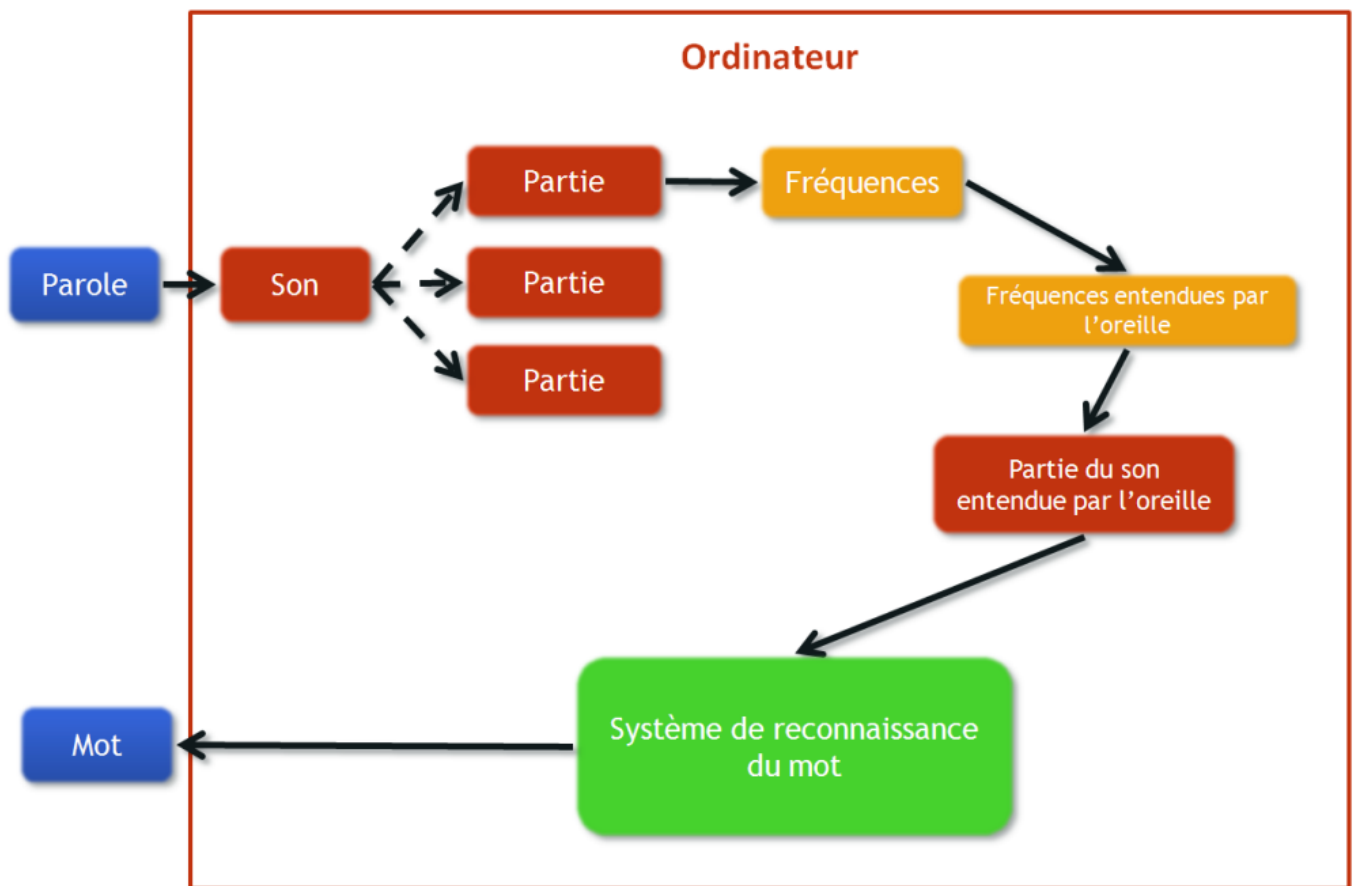


FIGURE 1.1 – Traitement du son pour le reconnaître

Utilisation de l'échelle de Mel Puisque le programme doit savoir *faire la différence entre des mots*, c'est-à-dire des sons identifiés tels quels par une oreille *humaine*, il faut donner au programme un comportement similaire à celui d'une oreille humaine. On utilise pour cela une échelle qui accentue certaines fréquences. En effet, il a été montré[5] (et ensuite appliqué [6]) que l'oreille ne perçoit pas toutes les fréquences de la même façon.

2. Analyse, formatage du signal

2.1 Introduction

Comme nous l'avons mentionné, même le plus élémentaire des systèmes de reconnaissance vocale utilise des algorithmes au carrefour d'une grande diversité de disciplines : reconnaissance de motifs statistiques, théorie de l'information, traitement du signal, analyse combinatoire, linguistique entre autres, le dénominateur commun étant le traitement du signal qui transforme l'onde acoustique de la parole en une représentation paramétrique plus apte à l'analyse automatisée. Le principe est simple : garder les traits distinctifs du signal et éviter au maximum de tout ce qui pourra en parasiter l'étude. Cette conversion ne se fait donc pas sans perte d'information, et la délicatesse de la discipline tient en la sélection judicieuse des outils les plus adaptés afin de trouver le meilleur compromis entre perte d'information et représentation fidèle du signal.

2.2 Prérequis

2.2.1 Qu'est-ce que le son ?

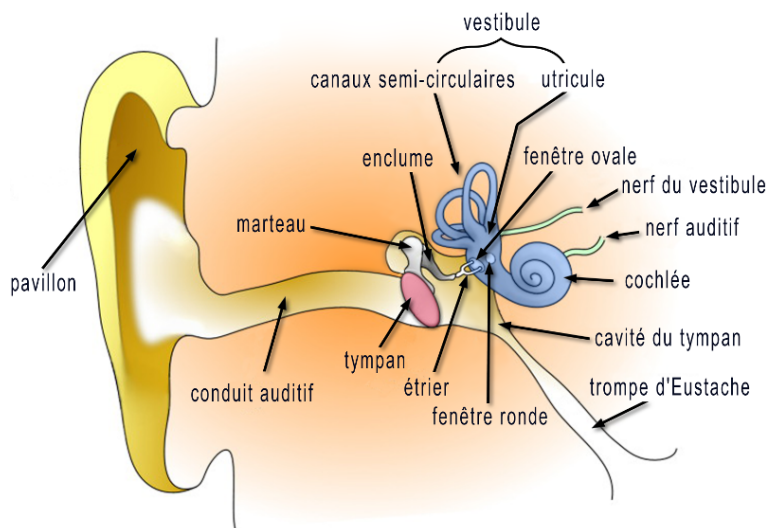


FIGURE 2.1 – Oreille humaine

Le son est une onde mécanique se traduisant par une variation de la pression au cours du temps. Cette onde est caractérisée par différents facteurs comme son amplitude à chaque instant, qui est en d'autres termes la valeur de la dépression à cet instant, et par les fréquences qui la composent et qui changent au cours du temps.

2.2.2 Comment le son est-il représenté dans l'ordinateur ?

En se propageant, l'onde mécanique qu'est le son fait vibrer la membrane du micro. L'amplitude de la vibration dépend directement de l'amplitude du son. La position de la membrane est enregistrée à

intervalles de temps réguliers définis par l'échantillonnage. L'échantillonnage correspond au nombre de valeurs prélevées en une seconde (principe [7]). Par exemple un échantillonnage à 44100 Hz correspond à relever la position de la membrane 44100 fois par secondes. La valeur de la position de la membrane est alors enregistrée sous la forme d'un entier signé codé sur n bits (n valant généralement 8,16,32 ou 64). Plus n est grand, plus la position de la membrane sera représentée de manière précise, et donc plus la qualité du son sera bonne. Grâce à l'échantillonnage et à la position de la membrane (n), on définit aisément le *bitrate*, qui correspond au débit d'information par seconde, de la façon suivante : $bitrate = n * \text{échantillonnage}$. Ce dont nous disposons donc pour analyser un signal, est la donnée de l'amplitude en fonction du temps la caractérisant.

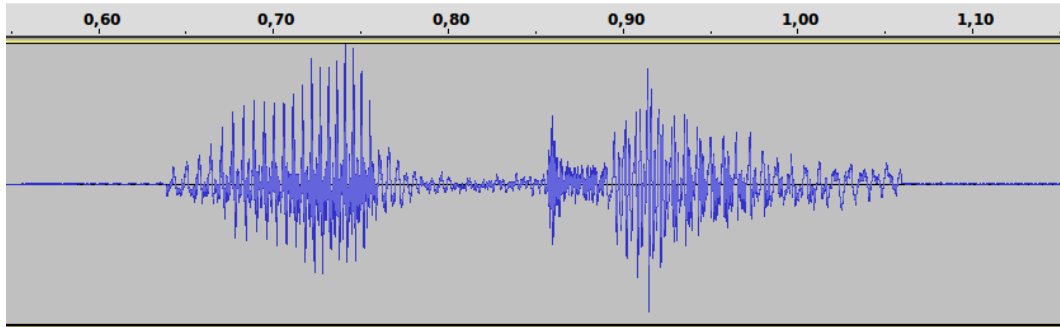


FIGURE 2.2 – Exemple audiogramme prononciation du mot "VICA"

2.3 Enregistrement, recadrage, filtrage HF

2.3.1 Synchronisation

Afin de synchroniser le début des enregistrements d'un mot, et de leur donner la même durée, nous avons eu l'idée de détecter les silences avant et après le mot pour les couper. Le signal est lissé à l'aide d'une moyenne sur plusieurs échantillons pour que les fluctuations inhérentes à l'enregistrement ne gênent pas notre fonction. On détecte alors le moment où le signal (en valeur absolue) dépasse pour la première fois une valeur seuil et celui à partir duquel le signal ne dépasse plus celle-ci. On sait alors où couper le signal d'origine, en élargissant légèrement la coupe afin d'éviter de supprimer des consonnes peu sonores. Cela permet en plus d'afficher un message d'erreur suspectant un enregistrement ayant commencé trop tard ou fini trop tôt. Deux problèmes se posent : en pratique, un bruit trop important perturbe le signal et le mot n'est plus détectable par l'amplitude des oscillations. Toutefois, pour l'enregistrement de notre base de données, une pièce calme et un micro de bonne qualité nous ont permis un découpage satisfaisant, ce qui ne résout pas définitivement le problème, l'utilisateur ne pouvant pas toujours se placer dans ces conditions, le signal est traité par un filtre anti-bruit.

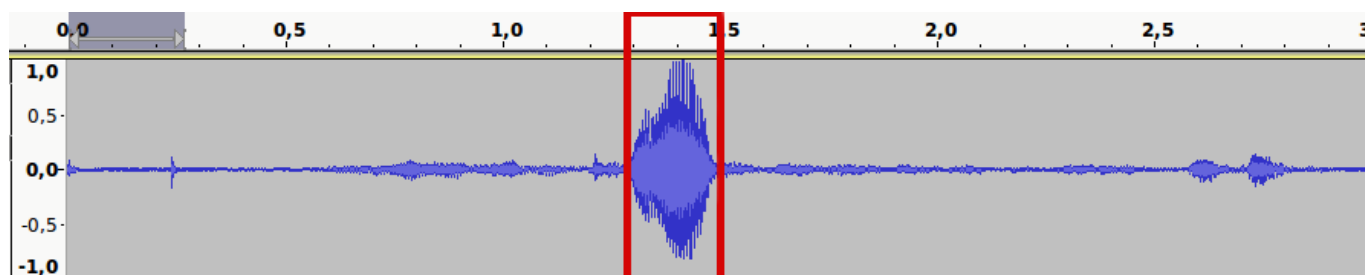


FIGURE 2.3 – Ensemble du son enregistré, la partie nous intéressant (le mot) est encadrée en rouge

Ce filtre consiste en l'utilisation de bibliothèques, SoX et ffmpeg([8] et [9]), qui permettent par l'étude d'un court laps de temps de bruit de soustraire le bruit de l'enregistrement. Nous n'avons pas cherché à traiter nous-même le bruit car il s'agit d'un problème complètement à part et qui ne demande pas les mêmes compétences que le traitement du signal effectué jusque là.

De plus, il a fallu déterminer la valeur de nos constantes de découpe (coefficient de lissage, coefficient de coupe, intervalle de temps de sécurité), qui dépendent bien sûr les uns des autres. Ceci a été fait de manière empirique sur plusieurs enregistrements de mots différents, permettant une découpe automatique la plus satisfaisante possible pour l'ensemble des mots.

2.3.2 Filtrage passe-haut

Les performances de tout système de reconnaissances dépendent fortement de la variabilité des données (locuteur, environnement, bruit, réverbération, ...). Plus ces données sont variables, plus le taux d'erreur sera grand et un système de reconnaissance qui se veut être utilisable dans la vie de tous les jours : (voiture, endroits bruyants) ; se doit d'y remédier. Ces effets se font particulièrement sentir dans les basses fréquences, c'est pourquoi le conditionnement du signal en vue de son étude comprend inmanquablement un filtre passe haut c'est-à-dire une accentuation de l'amplitude associée aux hautes fréquences et une diminution des basses fréquences. C'est le même principe qui est utilisé dans les égaliseurs des lecteurs de musique d'aujourd'hui qui proposent d'augmenter les basses ou les aigus. Les filtres passe-haut améliorent significativement les résultats de reconnaissance comme en témoignent les expériences de H.G. Hirsch P. Meyer et H.W. Ruehl dans leur papier[10].

Utiliser un filtre passe-haut présente comme avantage de ne pas nécessiter de procéder au préalable à une reconnaissance de silence contrairement aux techniques de réduction du bruit et de soustraction spectrale.

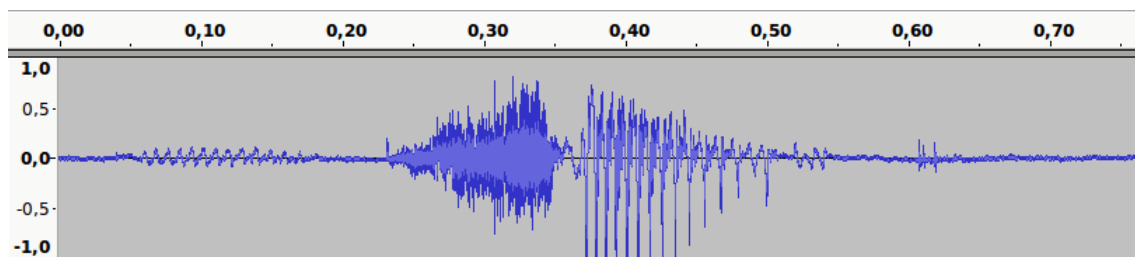


FIGURE 2.4 – Exemple d'un fichier son (représentant « Cinq ») **avant** application du filtre

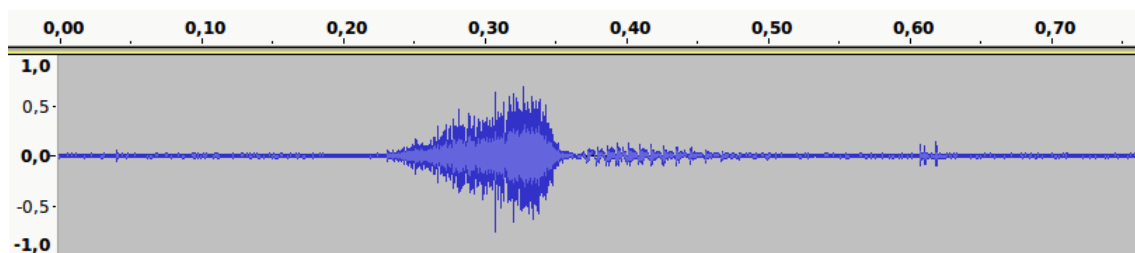


FIGURE 2.5 – Exemple du fichier son **après** application du filtre

Le signal étant caractérisé par une suite (x_n) d'amplitudes, comme présenté dans les prérequis, où n représente un instant de la musique déterminé par l'échantillonnage ; on opère linéairement la transformation suivante sur le signal : $y_0 = x_0$ et $y_n = x_n - 0.95 * x_{(n-1)}$ pour $n > 0$, où y représente le signal de sortie après transformation.

Cette opération consiste effectivement en un filtre passe-haut, en effet une telle formule part du principe que 95% d'un échantillon a pour origine l'échantillon précédent. Ce constat étant plus pertinent pour les hautes fréquences (car les pics de l'onde associée sont plus rapprochés et engendrent donc un pic d'amplitude plus régulièrement), l'influence des basses fréquences est donc discriminée.

2.4 Echantillonnage, fenêtrage

L'analyse du signal, pour accéder au domaine fréquentiel, s'affranchit de la dépendance temporelle. Le spectre obtenu ne correspond plus à une perception physique, mais à une moyenne temporelle du spectre perçu. Le procédé que nous avons mis en place pour pallier à ce problème est celui le plus couramment utilisé dans ce domaine : l'échantillonnage. Nous avons découpé le signal à traiter en petites séquences, qui, juxtaposées, approximent une échelle temporelle continue.

La taille des échantillons est un paramètre déterminant sur la qualité et la précision de l'analyse combinée finale. Une fois calculé, le spectre ne reflète plus du tout de dépendance temporelle. La durée d'un échantillon correspond ainsi à la durée minimale d'un événement sonore détectable. Il faut donc réduire cette durée autant que possible, pour obtenir une discrétisation temporelle le plus proche possible de la continuité. Il est en revanche nécessaire de conserver un certain nombre de points par échantillon. En effet, le spectre obtenu par l'analyse sera plus précis et proche de la réalité fréquentielle si le nombre de points du signal analysé est important. La meilleure technique pour contourner ce compromis est d'augmenter la fréquence d'échantillonnage. On obtient alors un nombre important de points qui s'étirent peu dans le temps.

Le théorème de Nyquist-Shannon[11] assure qu'un signal reproduit fidèlement toutes les fréquences inférieures à la moitié de sa fréquence d'échantillonnage. Une fréquence d'échantillonnage de 44100Hz (parfois 48000Hz) est donc suffisante pour couvrir la totalité d'une oreille humaine en bonne santé. L'utilisation la plus courante de l'enregistrement audio étant (à notre niveau) la restitution, le matériel et le logiciel à notre disposition se cantonnaient à ces fréquences d'échantillonnage. Nous avons ainsi dû trouver un compromis entre résolution fréquentielle et précision temporelle. L'hypothèse principale a été que les événements sonores et variations s'étalant sur une durée inférieure à 20 millisecondes n'étaient pas significatifs pour notre analyse. Le nombre de points a été par cette donnée, couplée à notre fréquence d'échantillonnage lors des enregistrements, à 44100Hz.

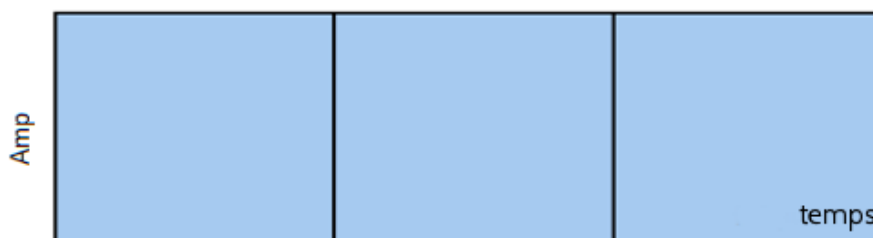


FIGURE 2.6 – Principe normal du fenêtrage



FIGURE 2.7 – Fenêtrage de Hann évitant les discontinuités

L'échantillonnage introduit par ailleurs des discontinuités aux bornes des morceaux, qui ne sont pas présentes dans le signal original. Le fenêtrage permet de réduire l'effet de ces discontinuités virtuelles. On découpe le signal en plus de morceaux, tout en conservant la même durée pour chaque échantillon. On obtient des "fenêtres", qui se recoupent les unes les autres. Pour que la même partie du signal ne soit pas retraitée à l'identique, on applique une fonction - dite fonction de fenêtrage, ou dans notre cas, fonction de Hann - qui diminue l'importance des valeurs situées aux extrémités de la fenêtre. Ce procédé a le désavantage de démultiplier le temps de calcul des étapes suivantes de l'algorithme (le nombre d'échantillons est bien plus important pour un signal de même longueur). Certaines applications (notamment pour les téléphones portables) devant réduire la complexité au maximum en font donc abstraction. Notre reconnaissance privilégiant plutôt la précision, et disposant d'une puissance de calcul largement suffisante pour conserver un rendu de l'ordre de la seconde, nous avons opté pour un fenêtrage important (recouvrement total d'un échantillon par ses voisins), au prix d'une multiplication du temps de calcul par deux.

2.5 Transformée de Fourier

Le domaine temporel est parfait pour l'acquisition et la restitution de l'audio, car il représente fidèlement la vibration de la membrane d'un micro ou d'une enceinte. L'oreille humaine base sa perception et sa reconnaissance sur le domaine fréquentiel. Il faut donc passer de l'un à l'autre, et ce grâce à l'utilisation de la transformation de Fourier. L'algorithme "intuitif" de calcul ayant, pour trouver le spectre d'un unique échantillon, une complexité en $O(N^2)$ (avec N le nombre de points par échantillons), il est nécessaire de trouver d'autres méthodes si l'on envisage des applications proches du temps réel. Heureusement, plusieurs approches se sont ouvertes à nous pour l'optimisation du temps de calcul.

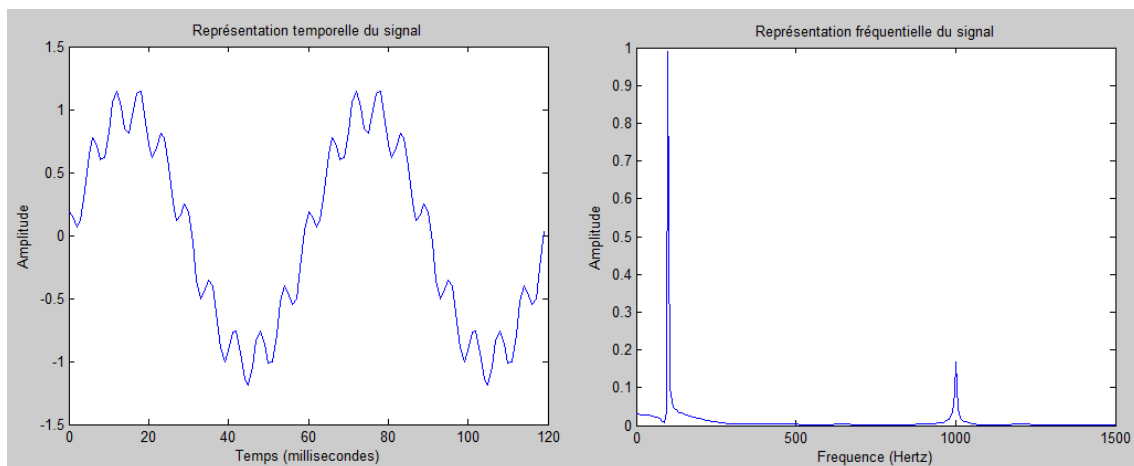


FIGURE 2.8 – Exemple de passage du domaine temporel (somme de cosinus) au domaine fréquentiel (pics pour les fréquences fondamentales)

Le calcul de la transformée de Fourier est incontournable en analyse du signal, et il a donné lieu à de nombreuses études. Des algorithmes optimisés pour diverses utilisations sont disponibles, et notre travail a surtout été d'identifier lequel s'adapterait à notre projet. La fonction que nous avons implémentée est l'algorithme de Cooley-Tukey, qui permet de réduire la complexité à $O(N \log_2(N))$, et qui repose sur le fonctionnement diviser pour régner. Le principe est dans un premier temps de diviser le signal à analyser en sous-tableaux de mêmes tailles, de manière croisée (par exemple deux sous-tableaux, pour les indices pairs et impairs). On calcule ensuite les transformées de Fourier de ces sous-tableaux, en opérant récursivement, jusqu'à obtenir des sous-tableaux dont la taille est un entier. On calcule leur transformée de Fourier, et on recombine les résultats obtenus. Cette méthode a l'avantage de pouvoir être couplée à d'autres algorithmes pour calculer les spectres des sous-tableaux dont la taille n'est pas un produit d'entier. Le meilleur cas est alors instinctivement un signal initial dont la longueur est une

puissance de deux. Il est même intéressant d'utiliser la technique du bourrage de zéros (zero padding), qui consiste à rajouter des zéros à la suite du signal pour atteindre la puissance de deux la plus proche. Cela ne change pas le spectre obtenu et augmente les performances. Dans notre cas, nous avons eu la possibilité d'ajuster la taille des échantillons. Nous avons ainsi choisi des échantillons de 1024 points, ce qui correspond, avec notre fréquence d'échantillonnage de 44100Hz, à une durée d'environ 23ms. Seul le dernier échantillon du signal est complété par des zéros.

De plus, comme les données sur lesquelles nous travaillons sont réelles, et que les calculs de la Transformée de Fourier Rapide (Fast Fourier Transform, ou FFT) s'effectuent avec des complexes, la première idée d'optimisation que nous avons eue est de calculer le spectre de deux échantillons à la fois, en créant des complexes à partir des deux signaux réels (l'un représente la partie réelle, l'autre imaginaire). On obtient rapidement les coefficients respectifs des deux échantillons par une simple opération sur le spectre résultant. Cependant, cette méthode ne divise le temps de calcul que par deux, et notre FFT demeure trop lente (plusieurs secondes pour un signal d'environ une seconde), surtout au regard du temps de calcul total de la reconnaissance en elle-même. La deuxième optimisation que nous avons donc appliquée est de passer le code de Python à C++, langage compilé beaucoup plus rapide. De plus, nous avons repensé les fonctions, de façon à éviter les appels récursifs. En effet, le travail sur des tableaux force une recopie à chaque appel de fonction, ce qui démultiplie la complexité du calcul. Le résultat est un algorithme qui s'effectue en moins d'une seconde, et qui peut s'inscrire dans un contexte d'exploitation en temps réel.

2.6 Simulation du comportement de l'oreille humaine

Des études de psycho acoustique ont montré que l'oreille humaine ne percevait pas les fréquences selon une échelle linéaire[5]. Il a donc été utile de définir une nouvelle échelle plus subjective : à chaque fréquence f , exprimée en Hertz, on fait correspondre une nouvelle fréquence selon une fonction censée représenter le comportement de l'oreille humaine. Par convention, la fréquence de 1000 Hz correspond à 1000 mel. Les autres fréquences mel sont ajustées de façon à ce qu'une augmentation de la fréquence mel corresponde à la même augmentation de la tonalité perçue. Cela conduit à la fonction *mel* suivante :

$$mel(f) = 2595 * \log(1 + f/700)$$

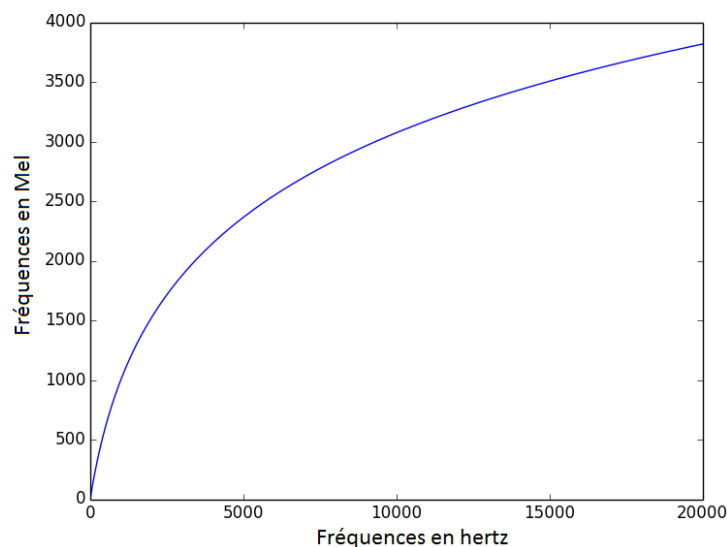


FIGURE 2.9 – Graphe de conversion

On remarque que le poids des hautes fréquences (supérieures à 1000 Hz) est diminué tandis que le poids des basses fréquences (inférieure à 1000 Hz) est augmenté.

Il est préférable d'employer cette échelle de fréquence dans l'algorithme de reconnaissance : ce dernier doit en effet différencier plusieurs mots selon la perception humaine, c'est-à-dire en simulant le comportement de l'oreille humaine.

2.7 Transformée inverse ou Décomposition en cosinus inverse

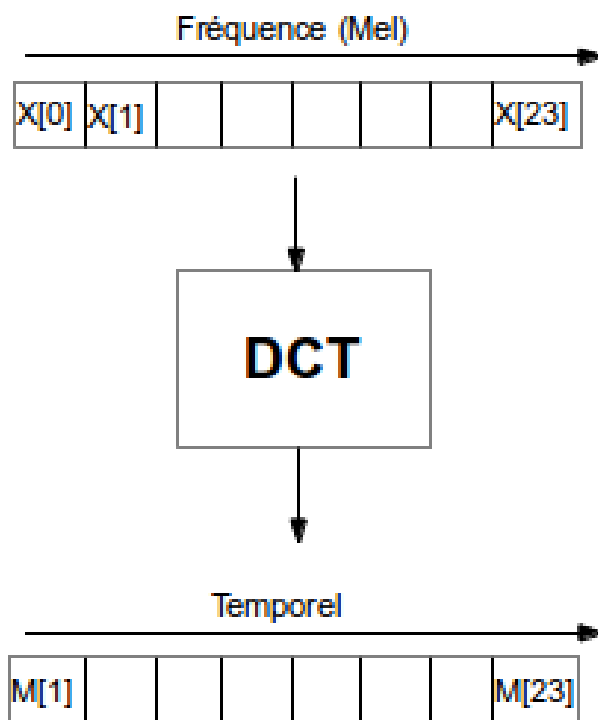


FIGURE 2.10 – Graphe de conversion

Dans les parties précédentes nous avons vu comment, à partir d'un extrait sonore échantillonné à 44100 Hz sur 16 bits, obtenir après transformée de Fourier et opérations sur le spectre, un tableau de 24 cases gradué en échelle Mel, représentant une fraction de l'extrait. Ce tableau exprimé ainsi en fréquences, pourrait a priori constituer une représentation satisfaisante de l'extrait sonore à l'instant considéré pour la suite de l'algorithme de reconnaissance, et servir à la comparaison avec le modèle au travers des chaînes de Markov cachées. Cependant ce n'est pas ce qui est fait, et l'on préférera une représentation en temporel de la fraction sonore considérée, ceci pour deux principales raisons.

- Opérer une transformation en cosinus inverse « décorrèle » les valeurs du tableau dans la mesure où dans une représentation en fréquentiel, les valeurs associées aux hautes fréquences sont très fortement corrélées avec ce qui se passe dans les basses fréquences. En effet, un signal sonore n'est jamais pur, c'est-à-dire constitué d'une seule fréquence, mais est un amalgame de signaux purs de fréquences multiples de celles d'autres signaux purs.
Le tableau qui sera traité par la suite grâce aux chaînes de Markov n'est plus constitué de 24 cases mais de 12 dont les 11 premières sont les premières cases du tableau obtenu après DCT. Si l'on tronquait le tableau avant d'opérer la DCT, on ne conserverait que l'information associée aux graves ce qui constituerait une perte trop importante de données.
- Ce retour au temporel se fait par la transformée en cosinus inverse. Il s'agit en terme simplistes

du pendant réel de la transformée de Fourier inverse, qui elle donne lieu à des coefficients complexes, lesquels dans le cadre d'une représentation temporelle n'ont que peu de sens. En termes plus mathématiques, la projection orthogonale du signal discret en fréquentiel ne se fait plus sur une base d'exponentielles complexes, mais de cosinus.

La DCT que nous avons utilisé, aussi connue sous le nom de DCTII se base sur la formule suivante :

$$M[k] = \sum_{n=0}^{B-1} (X[n] \times \cos(\pi \cdot k \cdot \frac{n+0.5}{B})) \times \sqrt{\frac{2}{B}}$$

avec B=24, X le tableau en échelle Mel, et M le tableau de sortie échelonné en temporel. D'autres formules équivalentes de DCT existent mais la DCTII est la plus largement répandue et utilisée.

3. Modélisation des mots à reconnaître par les modèles de Markov cachés

3.1 Objectifs

Le traitement effectué sur le son permet d'obtenir un tableau fréquentiel caractéristique du mot. Nous pouvons alors rentrer dans le vif du sujet et travailler sur la reconnaissance même des mots : il faut parvenir à comparer entre eux les empreintes ainsi obtenues. Pour cela, nous utilisons un système appelé les *modèles de Markov cachés*. Nous créons en fait un « graphe de sons » qui peut être parcouru à partir d'un mot. En appliquant un certain algorithme à un mot et à un graphe, on obtient une probabilité qui témoigne de l'adéquation du mot au graphe. Il nous suffit alors de déterminer quel graphe représente la plus grande probabilité ce qui nous donne la solution comme étant le *meilleur candidat*.

3.2 Prérequis et principe

Un modèle de Markov caché est un modèle statistique qui peut modéliser des processus physiques. Il fait appel aux structures d'automates[12].

3.2.1 Les automates

Un automate représente un système physique. Il est composé d'états (les cercles sur la figure), qui correspondent aux états du système réel, et de transitions (les flèches sur la figure), pour passer d'un état à l'autre. Il existe aussi la notion de chemin : par exemple pour passer de 0 à 3 sur la figure, il faut passer par 1 puis 2 : le chemin de 0 à 3 est 0,1,2,3.

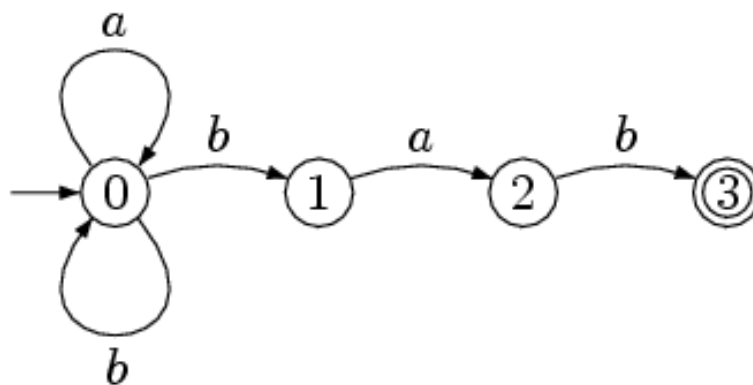


FIGURE 3.1 – Exemple d'automate « classique »

3.2.2 Les modèles de Markov cachés

Un modèle de Markov est un automate présentant *deux caractéristiques* en plus du principe de base d'un automate.

Tout d'abord, les transitions ne sont plus déterministes comme elles le sont dans le cas d'un automate mais sont **probabilistes**. Ainsi, il y a une certaine probabilité pour passer à chacun des autres état, une fois arrivé dans un état lors d'un parcours. C'est pourquoi le déplacement dans un automate ne dépend plus du passé du parcours mais uniquement de la position actuelle.

La seconde différence est que les données renvoyées lors du parcours d'un chemin de l'automate n'est plus la liste des états par lesquels passe le chemin : chaque état a maintenant des probabilités d'émettre certains signaux. On obtient donc une liste de signaux pour un chemin. Or, un état peut émettre plusieurs signaux différents et l'émission n'est que probabiliste ce qui rajoute au caractère non déterministe de l'automate.

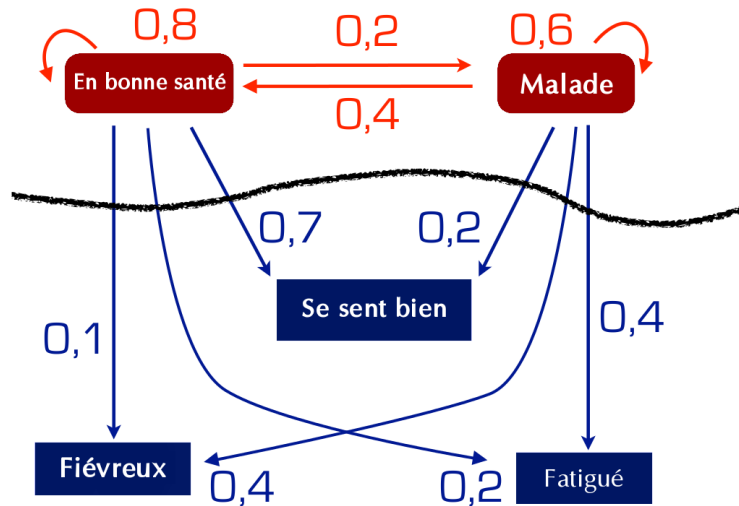


FIGURE 3.2 – Exemple d'un modèle de Markov caché

3.2.3 Application à la reconnaissance vocale

Les modèles de Markov cachés sont largement répandus dans la reconnaissance vocale ([13], [3] et [14]). Entre un modèle discret et un modèle continu, nous avons choisi ce dernier car les données en entrée ne font pas partie d'un ensemble fini : il existe une infinité de sons possibles pour un même phonème. Les modèles de Markov cachés sont particulièrement adaptés pour la reconnaissance vocale car ils permettent un apprentissage constant de la part du programme : celui-ci est capable d'apprendre de nouveaux mots de manière autonome, et de s'améliorer au-fur-et-à-mesure que la base de données de mots grandit.

Nous avons modélisé chaque mot par un automate, dont les états sont les différents phonèmes du mot. Lorsque l'on prononce un mot, on se dirige dans l'automate grâce aux phonèmes prononcés, jusqu'à rencontrer l'état final. Ceci permet de reconnaître le mot même si une syllabe dure plusieurs secondes : dans ce cas, on se contente de tourner en rond (en restant sur l'état 0 de la figure par exemple) dans l'automate jusqu'à rencontrer un nouveau phonème. Dans l'automate, la transition de l'état i à k représente la probabilité de passer de l'état i à k , c'est-à-dire la probabilité que le phonème $n^{\circ}k$ vienne tout de suite après le phonème $n^{\circ}i$.

3.3 Principaux algorithmes sur les modèles de Markov

Lorsque l'on fait passer un mot dans un automate, ie. qu'on s'oriente dans l'automate à l'aide des phonèmes, on peut calculer la probabilité que le mot corresponde à cet automate : on multiplie toutes

les probabilités rencontrées pendant le parcours. Elles dépendent bien sûr du chemin parcouru (i-e des transitions rencontrées). C'est le principe de l'algorithme *forward*.

L'algorithme de Baum-Welch permet d'optimiser un automate. En se plaçant dans l'ensemble des modèles de Markov, on cherche à faire converger une suite d'automates définis à l'aide de plusieurs versions d'un même mot vers un automate optimisé qui corresponde au mieux au mot.

3.4 Application à notre objectif

Résumons la situation lorsque l'on lance notre programme : d'un côté une base de données de mots, représentés chacun par un automate ; de l'autre, un fichier audio : le mot prononcé par l'utilisateur. Le programme se déplace dans chaque automate grâce au fichier audio, il s'oriente en fonction des phonèmes prononcés. Nous appellerons cette opération "faire passer un mot dans un automate".

L'algorithme *forward* permet donc de calculer la probabilité qu'un automate corresponde au mot prononcé : en comparant les probabilités dans chacun des automates, on sélectionne la plus grande et on a l'automate qui correspond le mieux au mot sélectionné.

L'algorithme de Baum-Welch permet l'apprentissage de nouveaux mots : pour chaque nouveau mot il crée un nouvel automate, et le rend le plus optimisé possible en s'appuyant sur la bibliothèque existante. C'est ce que fait la partie logicielle de notre programme, pour que les programmeurs puissent agrandir la base de données.

3.5 Phase d'apprentissage

Une fois l'algorithme de reconnaissance vocale implémenté, il nous a fallu l'améliorer. Deux aspects demandent un apprentissage de la part du programme. Il doit d'abord faire grossir l'ensemble des mots reconnus, de manière à pouvoir en reconnaître le plus possible. Mais il est aussi intéressant de lui faire apprendre un mot par des locuteurs différents. Plus le nombre de locuteurs est grand, plus l'algorithme peut être précis.

Enregistrer plusieurs personnes permet d'obtenir une diversité de spectres qui accroît la précision du programme.

Une fois un mot appris, il est également très utile qu'un même locuteur enregistre de nombreuses versions du mot. Nous avons fait pour notre locuteur 10 versions de chaque mot.

Pour mettre en place un apprentissage, nous avons des besoins matériels (stocker l'ensemble des mots reconnus) mais aussi des besoins humains, et en l'occurrence une diversité de voix.

Exemple:

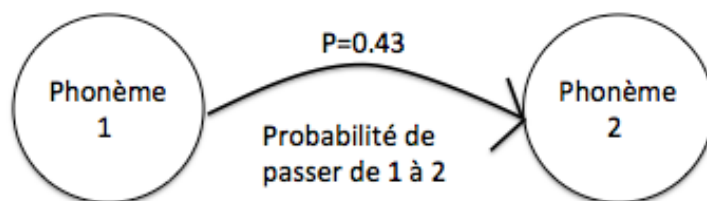


FIGURE 3.3 – Exemple de deux phonèmes et de la probabilité de passer du phonème 1 au phonème 2

3.6 Phase de reconnaissance

La phase de reconnaissance constitue le cœur du programme. Comme dit précédemment, le programme effectue l'algorithme *forward* sur chacun des automates et renvoie le mot le plus probable, après avoir comparé toutes les probabilités.

A l'origine, la phase de reconnaissance a été codée en Python. Cependant le temps d'exécution étant trop long, nous l'avons donc codé en C++, ce qui a permis de diviser le temps d'exécution par 50 000. Grâce à ce travail laborieux, le programme s'effectue en un temps proche de la seconde. Tout a été mis en place, notamment en amont avec le codage en C++ de la transformée de Fourier rapide, pour privilégier la rapidité de l'exécution.

Au départ nous n'avions qu'un seul locuteur pour faire la base de donnée des mots reconnus, ce qui ne permettait de faire fonctionner le programme que pour un seul utilisateur : celui qui avait enregistré les mots. Cependant nous avons enregistré plusieurs locuteurs, ce qui permet au programme de reconnaître plusieurs utilisateurs, même un utilisateur qui n'aurait pas encore enregistré de mot.

Récapitulatif

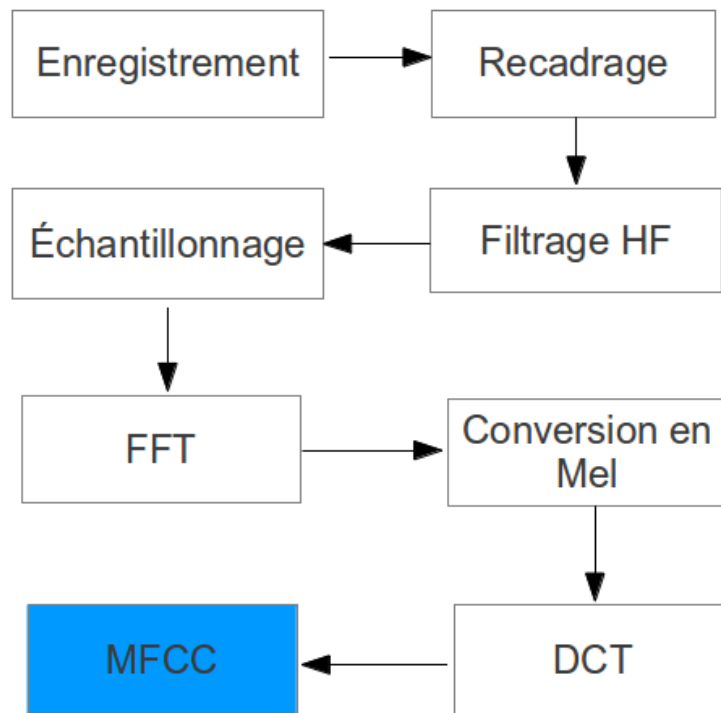


FIGURE 3.4 – exemple

Deuxième partie

Approche commerciale

1. Approche du développement du projet

Pour assurer une rentabilité à notre projet, il nous faut le penser, le structurer en vue d'une large distribution sous de multiples formes.

Nous sommes dotés d'une identité porteuse de ce projet. Le groupe de travail est baptisé The SpeechApp Company. Visuellement, elle se constitue en premier lieu d'un logo : Un micro, élément central du projet, dont la tête est le logo de Mines ParisTech, signe de notre appartenance à l'école et de l'aide qu'elle nous a apportée dans le projet.



FIGURE 1.1 – Logo de l'application

1.0.1 Choix d'une architecture optimale pour notre projet

Distribuer notre projet tel quel présenterait à ce stade de nombreux défauts :

- le coeur de notre technologie de reconnaissance vocale est directement accessible à tous.
- une interface unique en ligne de commande constitue un blocage majeur pour la majorité des utilisateurs finaux et empêche une intégration large à des applications tierces.

Étudions l'opportunité d'adopter une architecture client/serveur pour ce projet.

Dans ce scénario, divers clients logiciels, potentiellement indépendants de The SpeechApp Company pourraient communiquer par requêtes/réponses (spécifiées par une API¹) avec les serveurs de The SpeechApp Company. Ces derniers seuls auraient accès au coeur algorithmique du projet, qui resterait ainsi exclusivement entre nos mains. Par leurs requêtes, les clients demanderaient l'analyse automatique de mots, l'ajout de nouveaux mots ainsi que toute autre opération pertinente relative à l'analyse et la

1. API : interface de programmation permettant d'utiliser les fonctions proposés par le serveur

gestion d'une base de données de mots. L'accès à notre API serait monétisable forfaitairement ou à l'utilisation.

Les mots enregistrés par les clients seraient conservés dans des bases de données chez The SpeechApp Company. La location de ces bases de données hébergées serait monétisable. Alors, The SpeechApp Company pourrait prioritairement développer deux applications connectables au serveur : la première, SpeechCreator, permettrait l'enregistrement aisé de nouveaux mots dans les bases de données clients. La seconde, SpeechApp, permettrait, au travers d'une application Web riche, de tester la reconnaissance vocale en ligne.

Cette configuration permettrait aussi à une multitudes d'applications tierces d'utiliser notre technologie en ne voyant de l'extérieur qu'une API définissant le format des requêtes et réponses dans la communication entre clients logiciel et serveur.

Nous aboutirions alors à l'architecture représentée par le schéma suivant :

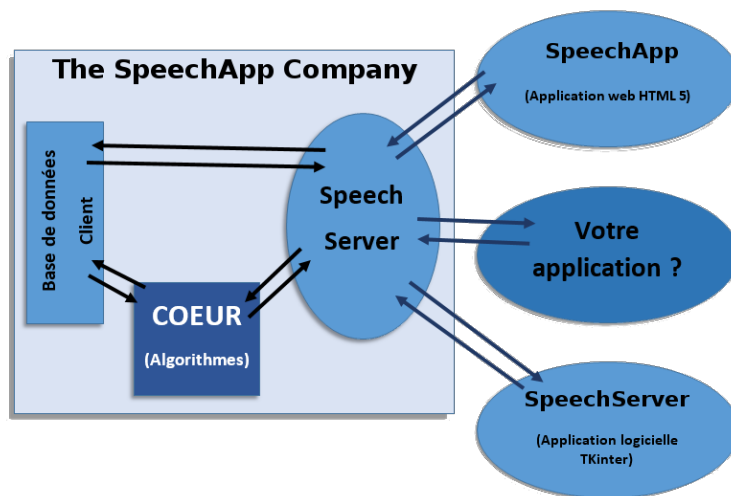


FIGURE 1.2 – Architecture proposée pour le projet The SpeechApp Company

Plus précisément dans le cadre des échanges entre le SpeechServer, les requêtes pourraient être traitées de la façon suivante : le client (au sens logiciel toujours) envoie au SpeechServer une requête HTTP POST contenant un formulaire avec en particulier son identifiant, son mot de passe, la base de données qu'il veut utiliser, l'action qu'il veut faire effectuer au SpeechServer, et les données d'entrée qui lui sont associées. La requête analysée par le SpeechServer, les opérations adéquates ayant été réalisées par le coeur algorithmique, le SpeechServer répond au client par une réponse HTTP POST² contenant des données au format XML³. Le client peut alors lire et interpréter la réponse donnée par le SpeechServer.

Avec ces spécifications, nous obtiendrions le cycle suivant pour la reconnaissance d'un mot par SpeechApp :

2. HTTP est un protocole de communication client-serveur développé pour web (Wikipédia)

3. XML est un langage informatique de balisage facilitant l'échange automatisé de contenus complexes (Wikipédia)

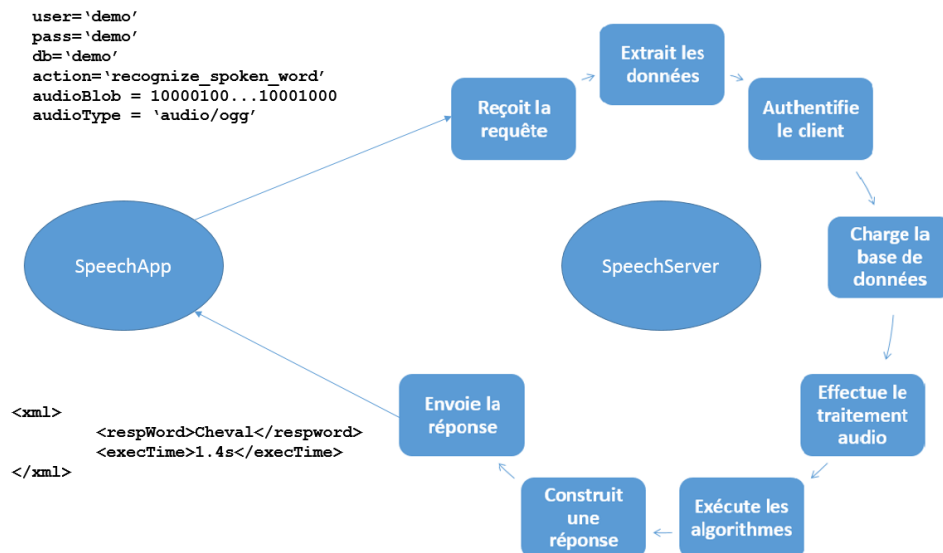


FIGURE 1.3 – Reconnaissance d’un mot par SpeechApp couplée au SpeechServer

L’architecture client/serveur proposée présenterait pour nous l’avantage de

- permettre la création d’un écosystème varié d’applications basées sur le coeur algorithmique de The SpeechApp Company via l’API de son SpeechServer, et générant ainsi des revenus
- conserver le coeur de notre travail entre nos mains et même de nous donner le contrôle sur toute la chaîne

L’architecture client/serveur proposée présenterait pour nos clients l’avantage de

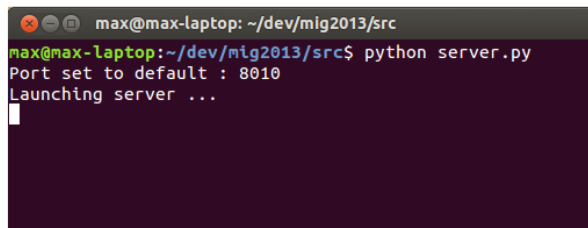
- ne pas se soucier du coeur algorithmique de la reconnaissance vocale, en n’y voyant que l’API de SpeechServer. Cette API peut offrir par ailleurs une grande liberté d’action
- n’avoir pas ou peu d’investissement initial de développement à effectuer, nos applications propriétaires SpeechApp et SpeechRecorder pouvant être intégrées sous forme de widgets aux applications tierces
- ne pas avoir à faire de lourds calculs eux-mêmes, ceux-ci étant réalisés par les machines de The SpeechApp Company
- les cycles de mises à jour seraient en majorité invisibles chez les clients, l’API restant immuables sur des cycles plus long (Long Term Support)

Au vu des nombreux avantages qu’elle présente, *nous avons donc opté pour une architecture modulaire client/serveur pour notre projet.*

1.0.2 Réalisation du SpeechServer

Le SpeechServer a été codé en Python. Python a une librairie standard suffisamment riche pour n’avoir à traiter ce problème qu’à un haut niveau (en réception de requêtes selon leurs méthodes). De plus, ce choix facilite les interactions avec le coeur algorithmique : des imports et appels de fonctions depuis le SpeechServer suffisent.

Finalement, le SpeechServer prend seulement la forme d’un programme Python à lancer sur un ordinateur.



```
max@max-laptop: ~/dev/mig2013/src
max@max-laptop:~/dev/mig2013/src$ python server.py
Port set to default : 8010
Launching server ...
```

FIGURE 1.4 – Le SpeechServer lancé

Il écoute alors les requêtes sur le port 8010 (par défaut) de l'ordinateur. Lorsqu'il en reçoit, il interagit avec le coeur algorithmique et le système de gestion de bases de données mis en place.

1.0.3 Système de Gestion de Base de Données (SGBD)

Le SGBD doit permettre de stocker et gérer les fichiers audio associés aux mots (au moins une dizaine d'enregistrements par mot), les modèles de markov cachés qui leurs sont associés ainsi que les données d'authentification des applications clientes.

Le standard actuel de gestion de bases de données est le modèle relationnel basé sur le langage SQL. Néanmoins, dans le cas précis de stockage de fichiers relativement lourds (> 0.1 Mo), la lecture/écriture des données directement sur le disque dur s'avère plus performante.

Nous avons donc fait le choix de stocker nos données sur le disque dur du serveur, en enregistrant les fichiers audio en format brut, et les autres données (modèles de Markov, données d'authentification) comme des objets Python, avec le module pickle de la librairie standard.

Un module python db.py a été developpé par nos soins pour gérer efficacement nos fichiers

Comme les accès en lecture/écriture à la mémoire RAM sont bien plus rapides que les accès aux disques durs, *on pourrait obtenir un gain de vitesse significatif pour la reconnaissance vocale en chargeant l'intégralité des données en mémoire RAM au démarrage du SpeechServer*. La vitesse en lecture/écriture sur un disque dur est de l'ordre de 50 Mo / s. Sur la RAM, elle est de l'ordre de 1 Go / s, soit un gain d'un facteur 20 pour les opérations en mémoire.

Néanmoins, la quantité de mémoire RAM nécessaire serait très importante, croissant linéairement avec le nombre de mots enregistrés. Les coûts engendrés pourraient être importants.

Tâchons de dimensionner l'infrastructure serveur dont nous aurions besoin.

1.1 Dimensionnement de l'infrastructure de calcul de The Speech App Company

Il nous faut d'abord définir les variables relatives au fonctionnement commercial de The Speech App Company ainsi que leurs valeurs de référence.

1.1.1 Hypothèses de fonctionnement

Soit M le nombre de clients de The SpeechApp Company. La référence sera $M = 1000$. Soit N le nombre moyen de mots dans les bases de données de chaque client. Nous prenons pour référence $N = 2000$ mots : un dictionnaire comme le Petit Robert en contient 60000.

Soit J le nombre de requêtes par seconde. Nous prendrons pour référence $J = 1000$ requêtes / s

On vise le traitement des requêtes en 1s. On gère donc J requêtes en simultané.

Les fichiers audio bruts envoyés par les clients au serveur pèsent environ 100 ko chacun. Les Modèles de Markov Cachés (MMC) associés aux mots pèsent environ 50 ko pour chaque mot. Lors des traitements sur ces fichiers audios, on estime qu'on a besoin de créer 5 fichiers audios temporaires, d'environ 100 ko chacun.

1.1.2 Dimensionnement en mémoire RAM et espace disque

Les fichiers audios bruts (10 par mot par défaut) et les MMC sont conservés sur le disque dur. Il faut donc $(10 * 100ko + 50ko) * N * M = 2.100To$ d'espace sur le disque dur.

Par ailleurs, on charge les MMC en mémoire, soit un espace RAM nécessaire de $50ko * N * M = 100Go$

Lors des opérations, si les $5 * 100 ko$ de fichiers temporaires sont créés en RAM, et qu'on gère environ $J = 1000$ requêtes en parallèle, il nous faut 500 Mo de RAM en plus, ce qui est marginal.

Au vu des capacités mémoire en informatique, toutes puissances de 2, *Dans le cadre de référence, il nous faut au moins 128 Go de RAM et 4 To d'espace disque*

1.1.3 Dimensionnement réseau

Pour la reconnaissance de mots, tâche la plus courante, Le serveur reçoit J requêtes de 100 ko (fichiers audio). Il faut donc recevoir 100 Mo / s de données. Le débit descendant (vers le serveur) doit donc être supérieur strictement à 100 Mo / s. Le serveur répond par des fichiers ne contenant que du texte, de taille négligeable devant celle des fichiers audio. Le débit montant (depuis le serveur) n'est donc pas un facteur discriminant dans le choix d'une connexion au réseau.

On veillera à avoir une connexion d'au moins 200 Mo / s)

1.1.4 Dimensionnement des éléments de calculs

Sur un ordinateur d'une puissance de calcul de 1 GFlops, on observe que lors de la reconnaissance d'un mot, l'unique opération dont la complexité dépend du nombre de mots en jeu, l'exécution de l'algorithme Forward (linéaire) prenait 0.005s sur une base de 100 mots.

Ainsi pour réaliser $J = 1000$ reconnaissances en simultannée sur des bases de $N = 1000$ mots avec un temps d'exécution de l'algorithme Forward de moins de 0.5s, *il nous faut une puissance de calcul d'au moins 100 Gflops.*

Les processeurs de dernière génération dédié au calcul atteignent ce niveau de performance. Le Intel Xeon E5-2670 atteint ainsi en théorie 330 Gflops

1.1.5 Choix de l'infrastructure et coûts liés

Connaissant les caractéristiques minimales du serveur : en termes d'espace RAM, disque dur, de connexion réseau et de puissance de calcul, nous pouvons choisir le serveur le plus adapté à nos besoins.

L'hébergeur OVH propose une gamme de serveurs de calcul pour les entreprises :

GAMME ENTERPRISE 2014				
Modèle	SP-64	SP-128	MG-128	MG-256
Prix	81.99€ HT /Mois	131.99€ HT /Mois	202.99€ HT /Mois	302.99€ HT /Mois
Installation	99.99€ HT	99.99€ HT	99.99€ HT	99.99€ HT
CPU	Intel Xeon E5-1620v2	Intel Xeon E5-1650v2	Intel Xeon 2x E5-2650v2	Intel Xeon 2x E5-2670v2
Cores / Threads	4c / 8t	6c / 12t	16c / 32t	20c / 40t
Fréquence / Burst	3.7 GHz+ / 3.9 GHz+	3.5 GHz+ / 3.9 GHz+	2.6 GHz+ / 3.4 GHz+	2.5 GHz+ / 3.3 GHz+
RAM	64 Go DDR3 ECC	128 Go DDR3 ECC	128 Go DDR3 ECC	256 Go DDR3 ECC
Disques Durs	2x 2 To SATA3 SSD ⁽¹⁾ / SAS ⁽¹⁾	2x 2 To SATA3 SSD ⁽¹⁾ / SAS ⁽¹⁾	2x 2 To SATA3 SSD ⁽¹⁾ / SAS ⁽¹⁾	2x 2 To SATA3 SSD ⁽¹⁾ / SAS ⁽¹⁾
RAID	SOFT HARD ⁽¹⁾	SOFT HARD ⁽¹⁾	SOFT HARD ⁽¹⁾	SOFT HARD ⁽¹⁾
Bande passante	300 Mbps ⁽¹⁾	300 Mbps ⁽¹⁾	400 Mbps ⁽¹⁾	500 Mbps ⁽¹⁾

FIGURE 1.5 – Gamme de serveur de calculs d'OVH

Nous devons disposer de 128 Go de RAM, de 4 To d'espace disque, de 100 Mo/s de connexion au réseau. Aussi le processeur Intel Xeon E5-1650 v2 ne dépasse pas les 70 Gflops et ne valide donc pas le critère de puissance de calcul. Le bi-ES-2650 atteint 140 Gflops, et le bi-ES-2670 330 Gflops.

Afin d'avoir une certaine marge, nous préférons prendre le processeur bi-Intel Xeon ES-2670. Nous sélectionnons donc le serveur MG-256 de OVH pour 303 euros HT / mois qui valide nos critères de performance avec une marge conséquente. À des fins de redondance, nous aurions besoin de 2 serveurs identiques, pour donc 606 euros HT / mois.

Nous nous sommes contentés d'un stockage des données intégralement sur disque dur et de moyens bien plus réduits lors de la phase de développement.

Les spécifications du SpeechServer et du SGBD ayant été définies, il devient possible et nécessaire de construire des applications se fondant dessus.

1.1.6 SpeechRecorder

Il est nécessaire de proposer aux clients une interface plus simple à appréhender que la console. C'est pourquoi nous avons développé l'application logiciel SpeechRecorder, qui permet aux clients enregistrés dans nos bases de données d'authentification (s'acquittant d'une licence), d'ajouter des mots à leurs bases de données. Elle devrait permettre à terme, de gérer l'intégralité des bases de données client.

Cette interface a été réalisée avec la librairie TKinter de Python, la librairie graphique Python la plus simple et la plus largement disponible : elle est incluse dans les paquetages de base de Python.

FIGURE 1.6 – Authentification d’un client au SpeechRecorder

FIGURE 1.7 – L’interface du SpeechRecorder

Pour entrer un nouveau mot dans une base de données client, par défaut 10 enregistrements sonores sont pris par SpeechRecorder. La librairie additionnelle pyaudio est utilisée pour ce faire.

1.1.7 SpeechApp

SpeechApp est le démonstrateur principal de notre projet. Il s’agit d’une application web permettant au grand public de tester notre technologie de reconnaissance vocale de mots isolés. À l’aide des dernières APIs HTML5 (élaborées depuis le début d’année 2013), l’utilisateur peut s’enregistrer sans l’installation de logiciel auxiliaire. Son enregistrement audio est transmis au SpeechServer (selon le schéma spécifié plus haut) qui renvoie le mot trouvé.

Pour ce démonstrateur, une application web a été choisie car elle fonctionne sur tout terminal doté d’un navigateur web récent sans nécessiter la moindre installation : nous l’avons conçue de façon à ce qu’elle soit adaptée aussi bien aux grands écrans d’ordinateurs, qu’à ceux plus petits des tablettes et smartphones. On qualifie ce type de design de "Responsive".

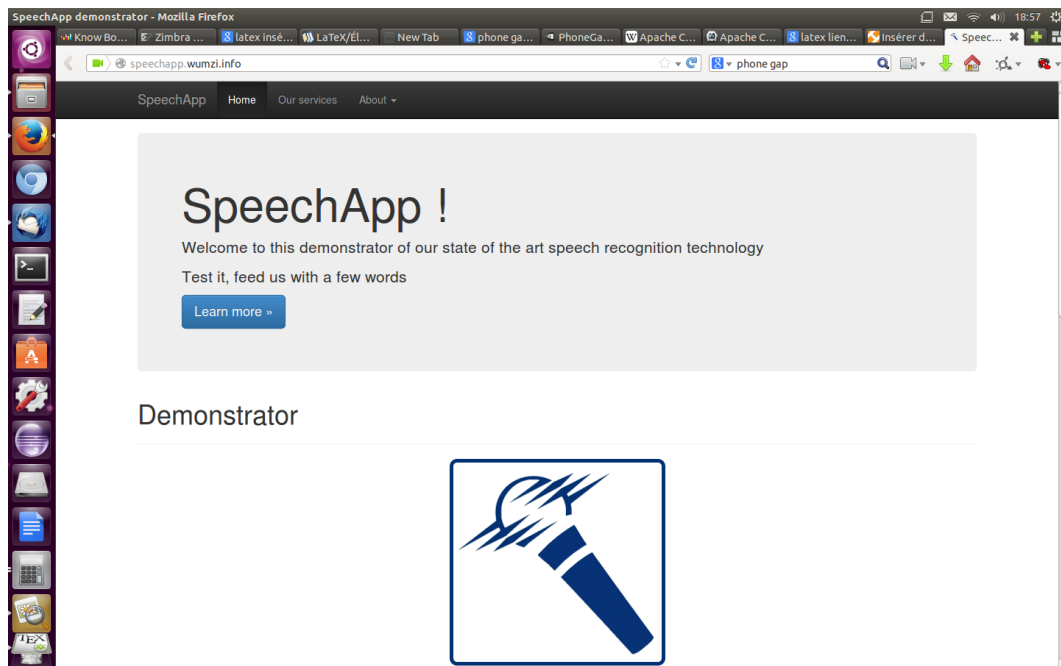


FIGURE 1.8 – Le démonstrateur SpeechApp sur ordinateur

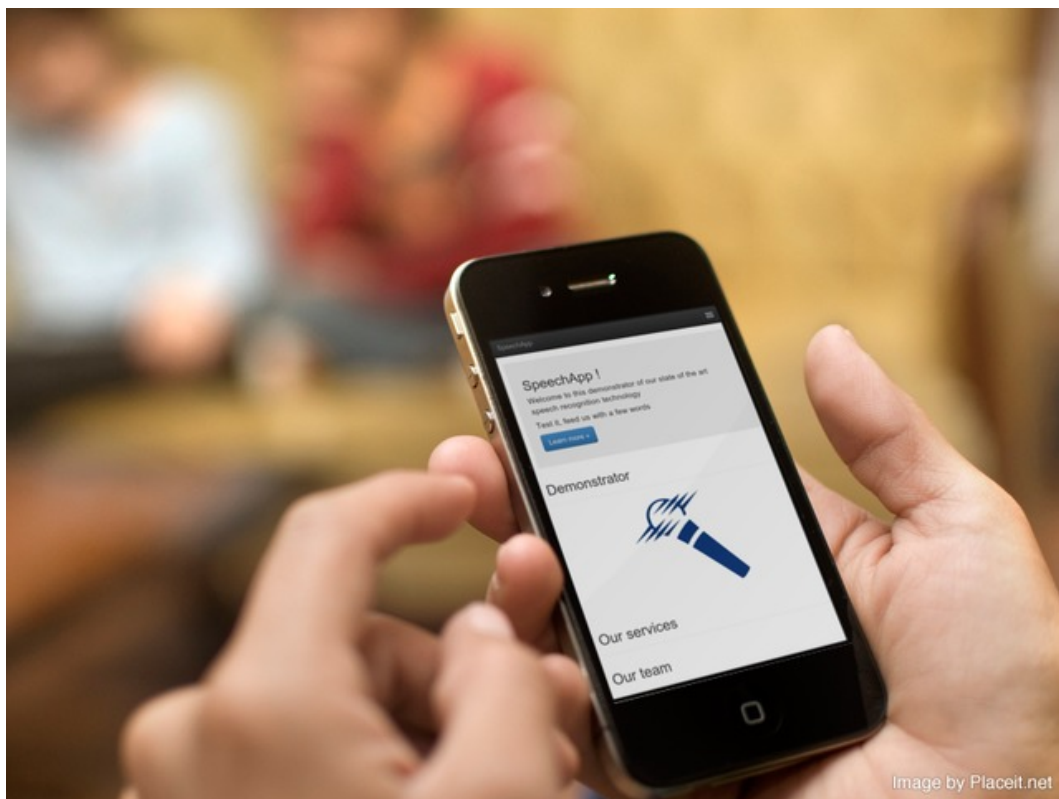


FIGURE 1.9 – Le démonstrateur SpeechApp sur iPhone



FIGURE 1.10 – Le démonstrateur SpeechApp sur iPad

Une difficulté néanmoins a été de rendre l'enregistrement audio fonctionnel sur la majorité des navigateurs. Nous assurons la compatibilité pour les moteurs Gecko et WebKit récents soit les dernières versions de Firefox, Chrome et Safari ainsi que leurs éditions mobiles. s

SpeechApp n'a en elle-même pas d'autre fonction que celle de démonstrateur, néanmoins ses modules peuvent être distribués aisément, sous forme de widgets intégrables n'importe où.

De plus, une application web reprenant des modules de SpeechApp pourrait être transformée aisément en application native pour smartphone iOS, Android, Windows Phone, Firefox Mobile ou encore en application Windows 8. Des frameworks open-source font ce travail presque automatiquement (par exemple <http://phonegap.com/>).

2. Applications

La reconnaissance vocale est une technologie promise à un futur radieux ; les plus grands noms de l'informatique, dont Bill Gates, annonçaient il y a quelques années qu'elle allait remplacer les claviers d'ici peu. Il s'avère aujourd'hui que leurs prédictions ne sont pas encore réalisées, il est tout à fait possible qu'elles se réalisent plus tard que prévu. Le principal obstacle à l'explosion de cette technologie étant le manque de fiabilité total, mais avec les progrès à venir, la technologie deviendra de plus en plus sûre.

L'armée étatsunienne a bien compris le potentiel de cette technologie : elle investit massivement depuis des années dans la recherche pour la développer[15]. Elle est d'ailleurs déjà utilisée sur certains avions de chasse, et pas seulement aux Etats-Unis : en France, en Angleterre et en Suède aussi notamment. Vu les investissements massifs, il y a fort à penser que les armées de ces différents pays ont des techniques bien plus avancées que celles connues du grand public, qui sont déjà plutôt performantes. Pour le moment, les commandes vocales ne servent pas encore à des fonctions critiques comme lancer un missile, et elles demandent toujours la confirmation du pilote avant d'exécuter une action. Elles libèrent néanmoins considérablement le pilote de beaucoup de tâches secondaires, ce qui lui permet de se concentrer sur les fonctions critiques. La technologie est également utilisée sur certains hélicoptères. Dans les deux cas, elle demande une grande fiabilité dans des conditions de stress et de bruit ambiant énorme (en particulier pour les hélicoptères, dans lesquels les pilotes n'ont souvent pas de casque anti bruit). Dans ce domaine, les perspectives sont donc très intéressantes financièrement mais elles demandent un savoir-faire qui est totalement hors de notre portée.

La reconnaissance vocale est également utilisée dans le contrôle aérien[16], et pourrait à terme remplacer les contrôleurs aériens. En effet, les phrases utilisées dans ce contexte sont très typées, ce qui favorise la reconnaissance (phrases souvent identiques, syntaxe très simple, prononciation très articulée). La technologie est donc moins avancée que dans le domaine de l'armée, et elle est déjà utilisée aux Etats-Unis, en Australie, en Italie, au Brésil et au Canada. Notre produit pourrait servir à ce type d'application, en créant une base de données spécifique au contrôle aérien.

La reconnaissance vocale se développe dans de nombreux domaines professionnels où les tâches administratives prennent beaucoup de temps, notamment la médecine, le droit et la police. En médecine[17], elle permet de remplir des rapports médicaux automatiquement : une simple relecture est alors nécessaire. Elle est notamment déjà utilisée dans 95% des hôpitaux aux Pays-Bas. Pour le droit, elle pourrait remplacer le travail du greffier pour prendre des notes dans les tribunaux. Et pour la police[18], elle permet de rédiger des rapports environ trois fois plus vite qu'au clavier. Le besoin de fiabilité est bien moindre dans ces domaines que dans les domaines de l'armée ou du contrôle aérien, une relecture est souvent largement nécessaire. Dans le domaine du droit, il faut néanmoins prendre en compte les conditions particulières d'enregistrement (brouhaha ambiant, émotions dans la voix, volume variable...). Notre produit peut tout à fait servir à ce type d'applications, à condition de créer une base de données spécifique aux domaines concernés.

Une autre application possible de la reconnaissance vocale est l'aide aux handicapés[19], par exemple des commandes vocales pour une chaise roulante. Les phrases utilisées sont très typées (avancer, reculer,...) donc la technologie n'a pas besoin d'être très avancée. De plus, avec la possibilité qu'offre

notre produit d'ajouter ses propres mots à la base de données, l'utilisateur lui-même peut rentrer les commandes ce qui assure un taux de reconnaissance très élevé. Notre produit peut donc bien s'adapter à cette utilisation.

La technologie est également très utilisée pour un usage plus ludique : fonctions de recherche dans les téléphones mobiles, les ordinateurs, robotique, jeux vidéo, traduction automatique,... Notre produit, dans sa version pour les particuliers, peut servir à ces usages même si la concurrence ne manque pas.

Enfin, la reconnaissance vocale peut servir à des fins sécuritaires, pour des vérifications d'identité. Il s'agit alors de reconnaître le locuteur, ce que notre produit ne permet pas.

Pour conclure, les applications pour notre produit sont assez nombreuses, et la demande est de plus en plus forte, ce qui montre sa pertinence.

3. Budget, modèle économique

3.1 Introduction

Après les études techniques et théoriques, l'étude économique est une nécessité. Elle est au coeur des problématiques de l'ingénieur, car c'est elle qui permet de dire si le projet est viable ou non. Dans le cas de la programmation d'un logiciel de reconnaissance vocale, divers facteurs sont à prendre en compte, comme les salaires des employés, la communication sur le produit ou les impôts à payer. Il s'agit également de trouver le meilleur moyen pour vendre le logiciel. Faut-il le vendre pour iPhone sur l'App Store ? Le réserver à un public restreint (majoritairement des entreprises) ou le proposer également à des particuliers ? La concurrence importante nous oblige à être à la fois ambitieux et prudent. Nous avons donc décidé d'envisager à la fois la vente sur notre site internet d'un logiciel pour les particuliers, et de proposer des licences en parallèle, permettant notamment aux entreprises d'accéder à nos bases de données, les compléter et créer leurs propres dictionnaires.

3.2 Les salaires

Treize employés travaillent sur le projet, pendant un temps effectif d'environ un mois. Parmi eux, un chargé des ressources humaines pour un salaire de 2750 € brut mensuel[20], un chargé d'étude de marchés, pour un salaire de 2700 € brut mensuel, les autres étant considérées comme des développeurs de moins de deux ans d'expérience, avec un salaire de 2290 € brut mensuel [21]. Sur ce salaire brut, l'employé paye environ 22% de charges salariales, et l'entreprise 44% de charges patronales.

SALAIRES

Catégorie	Salaire brut	Charges salariales	Salaire net	Charges patronales	Budget
Personnel					
David Vitoux	2290,00 €	503,80 €	1786,20 €	1007,60 €	3 297,60 €
Axel Goering	2290,00 €	503,80 €	1786,20 €	1007,60 €	3 297,60 €
Sofiane Mahiou	2290,00 €	503,80 €	1786,20 €	1007,60 €	3 297,60 €
Maxime Ernoult	2290,00 €	503,80 €	1786,20 €	1007,60 €	3 297,60 €
Adrien De La Vaissière	2290,00 €	503,80 €	1786,20 €	1007,60 €	3 297,60 €
Clément Joudet	2290,00 €	503,80 €	1786,20 €	1007,60 €	3 297,60 €
Clément Roig	2290,00 €	503,80 €	1786,20 €	1007,60 €	3 297,60 €
Anis Khlif	2290,00 €	503,80 €	1786,20 €	1007,60 €	3 297,60 €
Paul Mustière	2290,00 €	503,80 €	1786,20 €	1007,60 €	3 297,60 €
Matthieu Denoux	2290,00 €	503,80 €	1786,20 €	1007,60 €	3 297,60 €
Julien Caillard	2290,00 €	503,80 €	1786,20 €	1007,60 €	3 297,60 €
Nathanaël Kasriel	2750,00 €	605,00 €	2145,00 €	1210,00 €	3 960,00 €
Thomas Debarre	2750,00 €	605,00 €	2145,00 €	1210,00 €	3 960,00 €
Total	30690,00 €	6751,80 €	23938,20 €	13503,60 €	44193,60 €

FIGURE 3.1 – Salaires

3.3 Le compte de résultat prévisionnel

Le compte de résultat prévisionnel dresse l'ensemble des charges (fixes et variables) de l'entreprise, ainsi que ses produits (recettes). Pour parvenir à un équilibre budgétaire, il nous faut, pour la première année, vendre 26000 logiciels à un prix de 4,17 € hors taxes, et une dizaine de licences permettant d'accéder à nos bases de données pour un prix de 833,33 €. Au niveau des charges, l'ensemble des salaires cités plus haut est à prendre en compte, ainsi que le coût de notre campagne de publicité. Celle-ci peut être décrite en deux principaux pôles : des articles de journaux spécialisés, gratuits, et des annonces google. On peut estimer le prix d'une telle annonce à 10 centimes d'euros le clic. En estimant que 10% des visiteurs du site par l'intermédiaire de l'annonce vont acheter le produit, on peut évaluer le coût de la publicité à 26 000 €. Enfin, l'entreprise aura besoin, pour parvenir à fournir ses services de deux serveurs capable de traiter 1000 requêtes par seconde sur une base d'un million de mots pour un total de 606 € par mois. La différence des produits et des charges donne alors un chiffre de 62 710 €.

COMPTE DE RÉSULTAT PRÉVISIONNEL

	Produit					Charges	
	Vente	Prix unité (Hors taxes)	Prix unité TTC	Nombre	Total	Salaires	45 777,00 €
	Logiciel	4,17 €	5,00 €	24000	120 096,00 €	Frais pub (google)	24 000,00 €
	Licences	833,33 €	1 000,00 €	10	10 000,00 €		
Total	60319,00 €						

FIGURE 3.2 – Compte de résultat prévisionnel

3.4 Le bilan

Actifs		Passif	
Actifs incorporels	0,00 €	Fonds propres	0,00 €
Créances	0,00 €	Dettes long terme	100000,00 €
Actifs immobiliers	0,00 €	Compte de Résultat prévisionnel	60319,00 €
Créances clients	0,00 €		
Trésorerie	0,00 €		

FIGURE 3.3 – Bilan

Le bilan prend en compte l'actif et le passif de l'entreprise. Cette année, celle-ci n'a pas d'actif réel. Pas de trésorerie, de créances ou d'actifs immobiliers et incorporels. Son passif ne contient pas de fonds propres, et le compte de résultat prévisionnel a été explicité plus haut. On peut en revanche considérer que nous avons effectué un prêt à long terme de 100 000 €, afin de financer les prémices du projet.

3.5 Les impôts

S'agissant des impôts, nous devons dans un premier temps reverser à l'Etat la TVA sur les produits que nous vendons, à un taux de 20% à compter du 1er Janvier 2014. Le logiciel étant vendu 4,17 € et la licence 833,33 €, le total de la TVA à reverser sera de 27 020 €. Ensuite, l'impôt sur les sociétés est à un taux de 33% sur les bénéfices. A partir du bilan et de la TVA, on peut estimer nos bénéfices à 34 690 €, et donc un impôt sur les bénéfices à hauteur de 11 448 €[22].

Impôts		
Sur les sociétés	33% des bénéfices	11 318,93 €
TVA	20% sur les ventes	26 019,20 €

FIGURE 3.4 – Impôts

3.6 Conclusion et vue sur le long terme

En considérant un prêt à un taux de 3% sur cinq ans, et le prêt de locaux et ordinateurs par un incubateur (l'école des Mines par exemple) l'entreprise est viable la première année à partir de 25 000 téléchargements. En utilisant les mêmes calculs, pour les quatre années qui suivent, sans faire de mise à jour, il faudrait en moyenne 13 000 ventes de logiciels par an, et 3 ventes de licences.

Conclusion

Le marché de la reconnaissance vocale est pour le moment assez restreint, mais est appelé à grandir dans les prochaines années. Si les systèmes de reconnaissance vocale fleurissent sur les objets multimédias à usage personnel, comme les ordinateurs portables ou les téléphones mobiles, ils servent uniquement à simplifier un peu certaines tâches de l'utilisateur, et ne sont en pratique que très peu utilisés, ce qui s'explique par leurs performances moyennes. Le représentant le plus utilisé de ce type d'usage de la reconnaissance vocale est probablement Siri sur les téléphones mobiles iPhone d'Apple, mais il reste assez peu utilisé malgré la grande popularité de l'iPhone.

Dans le domaine des logiciels payants, pour un usage plus sérieux, le marché est dominé par les logiciels Dragon NaturallySpeaking de la firme américaine Nuance. Les prix, selon les modèles, varient entre environ 100\$ pour le modèle de base et environ 1000\$ pour les versions spécialisées dans un domaine professionnel. Le principe est que plus la base de données de mots est grande, plus les erreurs sont fréquentes ; Dragon NaturallySpeaking[23] propose donc des versions adaptées à un domaine particulier. Par exemple, il existe une version "juriste" avec une base de données contenant surtout du vocabulaire technique de droit, et une version "médecine" avec des termes techniques médicaux. Ces versions visant une cible très précise donc plus restreinte, ils sont vendus considérablement plus cher que les versions plus classiques. Cependant, la demande étant en constante augmentation - un tiers des radiologues français utilisent cette technologie, tout comme 95% des hôpitaux aux Pays Bas -, le marché est assez prometteur. En effet, cette technologie réduit considérablement les tâches administratives de ces professions : une simple relecture au plus est nécessaire. La réussite est renforcée par des taux de réussite exceptionnels avec des bases de données adaptées, et par l'absence de concurrence très forte.

Cependant, nous avons choisi de concevoir un logiciel avec une base de données moins spécialisée pour un usage personnel : en effet, dans le temps qui nous est imparti, créer des bases de données étudiées spécialement pour un certain domaine (droit, médecine) nous paraissait très compliqué. Il aurait fallu faire une étude linguistique très poussée pour construire la base de données, alors que nous avons concentré l'essentiel de nos efforts sur l'algorithme de reconnaissance lui-même. Notre produit est donc destiné à un usage plus ludique, ou du moins personnel. Notre cible est donc légèrement différente, puisque les professionnels intéressés par notre produit doivent construire eux-mêmes leur base de données spécifique à leur domaine. L'inconvénient de cette approche est le désagrément de devoir ajouter soi-même les mots, l'avantage étant que la reconnaissance sera plus fiable puisque la voix de l'utilisateur elle-même sert de comparateur, et elle permet d'avoir une base de données réellement personnalisée (celles de Dragon, bien que dédiées à un domaine, ne sont pas totalement personnelles). Le prix envisagé de la licence pour cette utilisation de notre produit est comparable (de l'ordre de 1000€) à celui des versions personnalisées de Dragon.

Nous prévoyons également de mettre en vente une version à usage personnel, sans possibilité d'ajouts de mots, au prix de 5€. Il est difficile de prévoir le potentiel de cette version, puisque les concurrents sont très nombreux, de qualité et de prix très variables.

Troisième partie

Code

A. Code Principal

A.1 shell.py

```
def main(verbose=True, action=-1, verboseUtime=True):
    db = Db("../db/", verbose=verbose)
    #HMMs = {}
    #db.addFile("hmmList.txt", HMMs)
    choice = -1
    while( not choice in range(1,8) ):
        try:
            if verboseUtime:
                choice = int(input("Que_voulez-vous_faire_\n1-Enregistrer
un_element\n
2-Realiser_l'analyse_d'un_mot\n3-Tester\n4-Afficher_rÃ©sultats_
intermÃ©diaires\n5-Gestion_des_fichiers_de_la_base_de_donnees\n\
6-Creation_d'un_HMM\n7-Gerer_les_HMM\n-----\n"))
            else:
                choice = 2
        except NameError:
            print "Ceci_n'est_pas_un_nombre_"
            #####
            ###          ENREGISTREMENT          ###
            #####
            if choice == 1:
                #RÃ©aliser un enregistrement
                recorder(db)

            #####
            ###          ANALYSE D'UN SON          ###
            #####
            elif choice == 2:
                fileOk = False
                while not fileOk:
                    #On choisit le dossier Ã© afficher
                    print "Voici_la_liste_des_mots_a_etudier:_"
                    dirList = db.printDirFiles("waves/")
                    dirChoice = -1
                    while( not dirChoice in range(len(dirList)) ):
                        try:
```

```

        dirChoice = int( input( "Choisissez_un_fichier_a_
                                traiter_et_entrez_son_numero:_ " ) )
    except NameError:
        print "Ceci_n'est_pas_un_nombre_"
    print "Dossier_choisi:_", dirList[ dirChoice ]
    fileOk = True
    numeroTraitement = 0
    filesList = db.printFilesList( dirList[ dirChoice ] )
    print filesList
    action = int( input( "A_partir_de_quelle_action_souhaitez-vous
                        _agir_?\n0-Tout\n1-Filtre_passe-haut\n2-Fenetre_de_Hann\n3
                        -Transformee_de_Fourier_Rapide\n4-Fonction_Mel\n5-Creation
                        _de_la_liste_Mel\n6-Transformee_de_Fourier_inverse\n7-
                        Creation_de_vecteurs\n_ " ) )
    for f in filesList:
        dirName = os.path.dirname(f)
        m = db.getWaveFile(f)
        if action == 2:
            content = db.getFile("handling/passe_haut_" + dirName
                                + "_" + str(numeroTraitement) + ".txt")
        elif action == 3:
            content = db.getFile("handling/hann_" + dirName + "_"
                                + str(numeroTraitement) + ".txt")
        elif action == 4:
            content = db.getFile("handling/fft_" + dirName + "_" +
                                str(numeroTraitement) + ".txt")
        elif action == 5:
            content = db.getFile("handling/mel_" + dirName + "_" +
                                str(numeroTraitement) + ".txt")
        elif action == 6:
            content = db.getFile("handling/mel_tab_" + dirName + "_"
                                + str(numeroTraitement) + ".txt")
        elif action == 7:
            content = db.getFile("handling/fft_inverse_" + dirName
                                + "_" + str(numeroTraitement) + ".txt")
        else:
            content = m[1]
        mot, log = handlingOneWord( content, db, dirName,
                                    numeroTraitement )
        if verbose:
            print log
        fileOk = False
        numeroTraitement+=1
        print "Mot_reconnu_pour_" + f + ":", mot

```

```

#####
###          TEST GLOBAL          ###
#####

```

```

elif choice == 3:
    fileName = recorder( db, "tmp", 1, False, 2, 1 )
    cutBeginning( Db.prefixPath + "waves/tmp/", fileName + ".wav", "
        cut_" )

```

```

syncFile( Db.prefixPath + "waves/tmp/", "cut_" + fileName + ".wav"
, "sync_" )
db.addFileToList("tmp/sync_cut_" + fileName + ".wav", "waves/")
finalTest("tmp/sync_cut_" + fileName + ".wav")

#####
###  RESULTATS INTERMEDIAIRES  ###
#####
elif choice == 4:
    print "Voici_la_liste_des_mots_a_etudier:_:"
    dirList = db.printDirFiles("storage/handling/")
    dirChoice = -1
    while( not dirChoice in range(len(dirList)) ):
        try:
            dirChoice = int( input( "Choisissez_un_fichier_a_traiter_
            et_entrez_son_numero:_:" ) )
        except NameError:
            print "Ceci_n'est_pas_un_nombre_!"
    print "Fichier_choisi:_:", dirList[dirChoice]
    amp = db.getFile("handling/" + str(dirChoice))
    db.addWaveFromAmp("output/" + str(dirChoice) + ".wav",44100,amp,"
    output/",False)

#####
###  GESTION DES FICHIERS DE LA BDD  ###
#####
elif choice == 5:
    choice3 = -1
    while( not choice3 in range(1,6) ):
        try:
            choice3 = int(input("Que_voulez-vous_faire_?\n1-Supprimer_
            un_fichier\n2-Supprimer_un_wav\n3-Synchroniser_la_BDD\
            n4-Synchroniser_les_wav\n5-Synchroniser_tous_les_
            fichiers\n"))
        except NameError:
            print "Ceci_n'est_pas_un_nombre_!"
    if choice3 == 1:
        print "Fichiers:_:"
        filesList = db.printDirFiles()
        dirName = "storage/"
    elif choice3 == 2:
        print "Dossiers_des_waves:_:"
        filesList = db.printDirFiles("waves/")
        dirName = "waves/"
    elif choice3 == 3:
        db.sync()
        db.sync("", "waves/")
    elif choice3 == 4:
        print "Voici_la_liste_des_mots_a_etudier:_:"
        dirList = db.printDirFiles("waves/")
        dirChoice = -1
        while( not dirChoice in range(len(dirList)) ):

```

```

        try:
            dirChoice = int( input( "Quel_mot_souhaitez_vous_
                                traiter?:_" ) )
        except NameError:
            print "Ceci_n'est_pas_un_nombre_"
    print "Dossier_choisi:_", dirList[ dirChoice ]
    filesList = db.printFilesList( dirList[ dirChoice ] )
    for f in filesList:
        cutBeginning( Db.prefixPath + "waves/", f, "" )
        syncFile( Db.prefixPath + "waves/", f, "" )
elif choice3 == 5:
    db.sync()
    db.sync("", "waves/")

#####
###      CREATION D'UN HMM      ###
#####
elif choice == 6:
    fileOk = False
    while not fileOk:
        #On choisit le dossier Ã afficher
        print "Voici_la_liste_des_mots_a_etudier:__"
        dirList = db.printDirFiles( "waves/" )
        dirChoice = -1
        while( not dirChoice in range(len(dirList)) ):
            try:
                dirChoice = int( input( "Choisissez_un_fichier_a_
                                        traiter_et_entrez_son_numero:__" ) )
            except NameError:
                print "Ceci_n'est_pas_un_nombre_"
        print "Dossier_choisi:_", dirList[ dirChoice ]
        fileOk = True
        filesList = db.printFilesList( dirList[ dirChoice ] )
        if len(filesList) < 6:
            print "Pas_assez_d'enregistrements"
            continue
        listVectors = []
        numeroTraitement = 0
        for f in filesList:
            dirName = os.path.dirname(f)
            m = db.getWaveFile(f)
            content, log = handlingRecording(m[1], db, dirName,
                                             numeroTraitement)
            listVectors.append(content)
            fileOk = False
            numeroTraitement+=1
        print "Sauvegarde:_:"
        db.addFile( dirList[ dirChoice ] + ".txt", listVectors, "hmm/" )
        hmmList = db.getFile( "hmmList.txt" )
        if hmmList.get( dirList[ dirChoice ] ):
            hmmList[ dirList[ dirChoice ] ].append( dirList[ dirChoice ] + ".
                txt" )
        else:

```

```

        hmmList[ dirList[ dirChoice ] ] = [ dirList[ dirChoice ] + ".txt"
        ]
    print "Extraction_:"
    db.addFile("hmmList.txt",hmmList)
    buildHMMs(hmmList.keys(),hmmList.values(), 500, Db.prefixPath
        + "hmm/")
    saveHMMs(Db.prefixPath + "hmm/save.hmm")
#####
###          LISTER LES HMMs          ###
#####
elif choice == 7:
    hmmList = db.getFile("hmmList.txt")
    print hmmList

def handlingOneWord(content,db,dirChoice,numeroTraitement,action=0):
    """ Fait le traitement d'un mot pour en construire les vecteurs de
        Markov et tester ensuite la compatibilit   avec les automates
        existants

        Retourne un tuple (motLePlusCompatible,log) """
    content,log = handlingRecording(content,db,dirChoice,numeroTraitement,
        action)
    loadHMMs(Db.prefixPath + "hmm/save.hmm")
    return recognize(content),log

def handlingRecording(content,db,dirChoice,numeroTraitement,action=0):
    log = ""
    if action <= 1:
        log += "Filtre_passe-haut_en_cours...\n"
        content = passe_haut(content)
        log += "Filtre_passe-haut_termine...\n"
        #db.addFile("handling/passe_haut_" + str(dirChoice) + "_" + str(
            numeroTraitement) + ".txt",content)
        #db.addWaveFromAmp("tmp/bob.wav", 44100, content)
        log += "Sauvegarde_effectuee...\n\n"
    if action <= 2:
        log += "Fen  tre_de_Hann_en_cours...\n"
        content = hann_window(content)
        log += "Fen  tre_de_Hann_terminee...\n"
        #db.addFile("handling/hann_" + str(dirChoice) + "_" + str(
            numeroTraitement) + ".txt",content)
        log += "Sauvegarde_effectuee...\n\n"
    if action <= 3:
        log += "Transformee_de_Fourier_rapide_en_cours...\n"
        content = fftListe(content,True)
        energyTable = construitTableauEnergy(content)
        for k in range(len(content)):
            for l in range(len(content[k])):
                content[k][l]=abs(content[k][l])
        log += "Transformee_de_Fourier_rapide_terminee...\n"
        #db.addFile("handling/fft_" + str(dirChoice) + "_" + str(
            numeroTraitement) + ".txt",content)
        log += "Sauvegarde_effectuee...\n\n"

```

```

"""
if action <= 4:
    log += "Application de la fonction Mel en cours..."
    for k in range(len(content)):
        content[k] = fct_mel_pas(content[k],10)
    log += "Application de la fonction Mel terminee..."
    db.addFile("handling/mel_" + str(dirChoice) + "_" + str(
        numeroTraitement) + ".txt",content)
    log += "Sauvegarde effectuee...\n"
if action <= 5:
    log += "Construction de la liste Mel en cours..."
    for k in range(len(content)):
        content[k] = mel_tab(content[k],10)
    log += "Construction de la liste Mel terminee..."
    db.addFile("handling/mel_tab_" + str(dirChoice) + "_" + str(
        numeroTraitement) + ".txt",content)
    log += "Sauvegarde effectuee...\n"
"""
if action <=5:
    log += "Application_de_la_fonction_Mel_en_cours...\n"
    for k in range(len(content)):
        content[k] = triangularFilter(content[k],RATE)
    log += "Application_de_la_fonction_Mel_terminee...\n"
    #db.addFile("handling/mel_" + str(dirChoice) + "_" + str(
        numeroTraitement) + ".txt",content)
    log += "Sauvegarde_effectuee...\n\n"
if action <= 6:
    log += "Transformee_de_Fourier_inverse_en_cours...\n"
    for k in range(len(content)):
        content[k] = inverseDCTII(content[k])
    log += "Transformee_de_Fourier_inverse_terminee...\n"
    #db.addFile("handling/fft_inverse_" + str(dirChoice) + "_" + str(
        numeroTraitement) + ".txt",content)
    log += "Sauvegarde_effectuee...\n"
if action <= 7:
    log += "Creation_de_vecteurs_HMM_en_cours...\n"
    content = creeVecteur(content, energyTable)
    log += "Creation_de_vecteurs_HMM_terminee...\n"
    #db.addFile("handling/vecteurs_" + str(dirChoice) + "_" + str(
        numeroTraitement) + ".txt",content)
    log += "Sauvegarde_effectuee...\n\n"
#db.logDump(str(dirChoice) + "_" + str(numeroTraitement),log)
#db.logDump(str(dirChoice) + "_" + str(numeroTraitement))
return content,log

#def handlingOneWord(content,db,dirChoice,numeroTraitement,action=0,
    hmmList=[]):
def finalTest(fileName = ""):
    db = Db("../db/",verbose=False)
    fileOk = False
    while not fileOk:
        #On choisit le dossier A~ afficher
        if fileName == "":

```

```

print "Voici_la_liste_des_mots_a_etudier:_:"
dirList = db.printDirFiles("waves/")
dirChoice = -1
while( not dirChoice in range(len(dirList)) ):
    try:
        dirChoice = int( input( "Choisissez_un_fichier_a_
                               traiter_et_entrez_son_numero:_:" ) )
    except NameError:
        print "Ceci_n'est_pas_un_nombre_"
print "Dossier_choisi:_:", dirList[dirChoice]
fileOk = True
numeroTraitement = 0
filesList = db.printFilesList(dirList[dirChoice])
print filesList
fileChoice = -1
while( not fileChoice in range(len(filesList)) ):
    try:
        fileChoice = int( input( "Choisissez_un_fichier_a_
                               traiter_et_entrez_son_numero:_:" ) )
    except NameError:
        print "Ceci_n'est_pas_un_nombre_"
print "Fichier_choisi:_:", filesList[fileChoice]
n = fileChoice
f = filesList[fileChoice]
d = dirList[dirChoice]
fileOk = False
else:
    f = fileName
    fileOk = True
    n = 1
    d = ""
dirName = os.path.dirname(f)
m = db.getWaveFile(f)
mot, log = handlingOneWord(m[1], db, d, n)
print "Le_mot_reconnu_est", mot
print "_____"
if __name__ == "__main__":
    main(False)

```

A.2 server.py

```

if __name__ == '__main__':
    import sys
    if len(sys.argv) >= 2:
        try:
            PORT = int(sys.argv[1])
        except TypeError:
            print("Please_provide_an_int_!")
    else:
        PORT = 8010
        print("Port_set_to_default:_:%s" % PORT)

```

```
print("Launching_server_...")
main.run(PORT)
```

A.3 gui.py

```
class Gui:
    def __init__(self):
        self.auth = AuthUser()
        #self.auth.login("giliam", self.auth.hashPass("test"))
        self.nbEnregistrement = 0
        self.listeEnregistrements = []
        self.db = Db("../db/")
        self.fenetre3enabled = False
        self.noiseOk = False

    def ouverture(self): #fonction qui ouvre une deuxi me
        fen tre graphique, et qui affiche le r sultat
        self.fenetre4=Tk()
        self.fenetre4.attributes('-alpha', 1) #plein  cran
        self.fenetre4.configure(background='white')
        self.fenetre4.title("MIG_SE_2013_-_Liste_des_mots_enregistr s")
        titre=Label(self.fenetre4, text='\nMIG_SE_2013',font=("DIN", "34",
            "bold"), fg='#006eb8', bg="#ffffff")
        titre.pack()
        titre_logiciel=Label(self.fenetre4, text="Reconnaissance_vocale\n\
            n\n\n",font =("DIN", "22"), bg="#ffffff")
        titre_logiciel.pack()

        panneau2=Label(self.fenetre4, text='Liste_des_mots_actuellements_
            reconnus:\n\n', font=("DIN", '14'), bg="#ffffff")
        hmmList = self.db.getFile("hmmList.txt")
        res = hmmList.keys()
        resultat=Label(self.fenetre4, text="\n".join(res),font =("DIN", "
            28", "bold"), fg="#16d924", bg="#ffffff")
        espace=Label(self.fenetre4, text="\n\n", bg="#ffffff")
        panneau2.pack()
        resultat.pack()
        espace.pack()
        bouton_fermer=Button(self.fenetre4, text='Quitter', command=self.
            fenetre4.destroy)
        bouton_fermer.pack()
        espace4=Label(self.fenetre4, text= '_\n_', bg="#ffffff")
        espace4.pack()
        self.fenetre4.mainloop()

    def creationHmm(self):
        self.bouton_enr.config(text="Terminer_l'enregistrement_du_HMM",
            command=self.fenetre3.destroy)
        self.noiseOk = False
        listVectors = []
        for l in self.listeEnregistrements:
            content = self.db.getWaveFile(l)
```



```

        content, log = handlingRecording(content[1], self.db, 0, 0, 0)
        listVectors.append(content)
hmmList = self.db.getFile("hmmList.txt")
if hmmList.get(self.mot):
    hmmList[self.mot].append("client_" + self.mot + ".txt")
else:
    hmmList[self.mot] = ["client_" + self.mot + ".txt"]
self.db.addFile("hmmList.txt", hmmList)
self.db.addFile("client_" + self.mot + ".txt", listVectors, "hmm/"
    )
buildHMMs(hmmList.keys(), hmmList.values(), 500, Db.prefixPath + "
    hmm/")
saveHMMs(Db.prefixPath + "hmm/save.hmm")

def enregistrer(self):
    if self.nbEnregistrement == 0:
        self.mot = self.saisirMot.get()
    if not self.noiseOk:
        self.errorMessage3.set("Vous n'avez pas encore enregistré le
            bruit!")
    elif self.mot == "":
        self.errorMessage3.set("Entrez un mot")
    else:
        self.errorMessage3.set("")
        fileName = recorder(self.db, "tmp", 1, False, 1, confirm=False,
            fileName=self.mot + "_" + str(self.nbEnregistrement))
        sox_handling(Db.prefixPath + "waves/tmp/" + self.mot + "_" +
            str(self.nbEnregistrement) + ".wav", Db.prefixPath + "
            waves/noise/" + self.mot + ".wav", Db.prefixPath + "waves/
            tmp/")
        cutBeginning(Db.prefixPath + "waves/tmp/", self.mot + "_" +
            str(self.nbEnregistrement) + ".wav", "cut_")
        syncFile(Db.prefixPath + "waves/tmp/", self.mot + "_" + str(
            self.nbEnregistrement) + ".wav", "sync_")
        self.listeEnregistrements.append("tmp/" + self.mot + "_" + str(
            self.nbEnregistrement) + ".wav")
        self.nbEnregistrement += 1
        self.bouton_enr.config(text="Lancer l'enregistrement numÃ©ro_
            " + str(self.nbEnregistrement + 1))
        if self.nbEnregistrement == NB_ITERATIONS:
            self.creationHmm()

def enregistrerNoise(self):
    self.mot = self.saisirMot.get()
    if self.mot == "":
        self.errorMessage3.set("Entrez un mot")
    else:
        self.errorMessage3.set("")
        fileName = recorder(self.db, "noise", 1, False, 1, confirm=False,
            fileName=self.mot)
        self.noiseOk = True

```

```

def loginAuth(self):
    loginIn = self.loginC.get()
    passwordIn = self.passwordC.get()
    if passwordIn == "" or loginIn == "":
        self.errorMessage.set("Il manque le pseudonyme ou le mot de passe!\n")
    else:
        self.auth.login(loginIn, self.auth.hashPass(passwordIn))
        self.fenetre2.destroy()
        self.bouton_loginpopup.config(text='Se dÃ©connecter')
        self.errorMessage.set("")
        Button(self.fenetre1, text='Liste des mots enregistrÃ©s',
                command=self.ouverture).pack()
        self.displayRecorder()

def registerAuth(self):
    loginIn = self.loginR.get()
    passwordIn = self.passwordR.get()
    if passwordIn == "" or loginIn == "":
        self.errorMessage.set("Il manque le pseudonyme ou le mot de passe!\n")
    else:
        if self.auth.getClient(loginIn) != "":
            self.auth.newClient(loginIn, self.auth.hashPass(passwordIn)), []
            self.errorMessage.set("Vous Ã©tes bien enregistrÃ©(e)\n")
        else:
            self.errorMessage.set("Le pseudonyme est dÃ©jÃ utilisÃ©")

def fenetre3destroy(self):
    self.fenetre3enabled = False
    self.fenetre3.destroy()

def displayRecorder(self):
    self.bouton_registerIn.pack()
    if not self.fenetre3enabled:
        self.fenetre3=Tk()
        self.fenetre3.attributes('-alpha', 1) #plein Ã©cran
        self.fenetre3.configure(background='white')
        self.fenetre3.title("MIG_SE_2013_-_Enregistrement")
        titre=Label(self.fenetre3, text='\nMIG_SE_2013', font=("DIN",
            "34", "bold"), fg='#006eb8', bg='#ffffff')
        titre.pack()
        titre_logiciel=Label(self.fenetre3, text="Reconnaissance_
            vocale\n\n\n\n", font=("DIN", "22"), bg='#ffffff')
        titre_logiciel.pack()
        self.errorMessage3 = StringVar(self.fenetre3)

        errorLegend=Label(self.fenetre3, textvariable=self.
            errorMessage3, fg='#B80002', bg='#ffffff')
        errorLegend.pack()

```

```

demande_mot=Label(self.fenetre3, text="Entrez le mot que vous
souhaitez enregistrer", font=("DIN", '14'),bg='#ffffff')
demande_mot.pack()
self.saisirMot=StringVar(self.fenetre3) # variable
pour recevoir le texte saisi
saisieMot=Entry(self.fenetre3, textvariable=self.saisirMot,
width=30,bg='#ffffff')
saisieMot.pack()
Label(self.fenetre3, text="Avant de procÃ©der aux
enregistrements, il convient d'enregistrer \
du bruit pour permettre efficacement le traitement du signal \n", font=(
"DIN", '14'),bg='#ffffff').pack()
self.bouton_bruit=Button(self.fenetre3, text='Enregistrer du
bruit', command=self.enregistrerNoise, fg='#ff0000') #
bouton qui enregistre et ouvre une nouvelle fenetre
self.bouton_bruit.pack()
Label(self.fenetre3, text="Il est nÃ©cessaire pour crÃ©er un
modÃ©le de Markov cachÃ© de procÃ©der Ã une dizaine d'
enregistrements du mÃªme mot\n\
Vous aurez deux secondes Ã chaque enregistrement pour prononcer votre
mot une fois l'acquisition lancÃ©e\n", font=("DIN", '14'),bg='#ffffff'
).pack()
espace1=Label(self.fenetre3, text= '\n',bg='#ffffff')
espace1.pack()
self.bouton_enr=Button(self.fenetre3, text='Lancer l'
enregistrement numÃ©ro 1', command=self.enregistrer, fg='#
ff0000') #bouton qui enregistre et ouvre une nouvelle
fenetre
self.bouton_enr.pack()
bouton_fermer=Button(self.fenetre3, text='Quitter', command=
self.fenetre3.destroy)
bouton_fermer.pack()
espace4=Label(self.fenetre3, text= '\n',bg='#ffffff')
espace4.pack()
self.fenetre3.enabled = True

def displayLogIn(self):

    if self.auth.connected:
        self.auth.logout()
        self.bouton_loginpopup.config(text='Se connecter / S\'inscrire
')
    else:
        self.fenetre2 = Tk()
        self.fenetre2.title("MIG_SE_2013")
        self.fenetre2.attributes('-alpha', 1)
        self.fenetre2.configure(background='white')

        self.loginC = StringVar(self.fenetre2)
        self.passwordC = StringVar(self.fenetre2)
        self.loginR = StringVar(self.fenetre2)
        self.passwordR = StringVar(self.fenetre2)
        self.errorMessage = StringVar(self.fenetre2)

```

```

errorLegend=Label(self.fenetre2, textvariable=self.
    errorMessage, fg='#B80002',bg='#ffffff')
errorLegend.pack()
Label(self.fenetre2, text="\nSe_connecter_\n", font=("DIN", '
    14'),bg='#ffffff').pack()

loginLabel=Label(self.fenetre2, text="Identifiant:", font=(
    "DIN", '14'),bg='#ffffff')
loginLabel.pack()
loginForm=Entry(self.fenetre2, textvariable=self.loginC, width
    =30)
loginForm.pack()

passwordLabel=Label(self.fenetre2, text="\nMot_de_passe:",
    font=("DIN", '14'),bg='#ffffff')
passwordLabel.pack()
passwordForm=Entry(self.fenetre2, textvariable=self.passwordC,
    width=30)
passwordForm.pack()

bouton_envoyer=Button(self.fenetre2, text='Se_connecter',
    command=self.loginAuth, fg='#000000')
bouton_envoyer.pack()

Label(self.fenetre2, text="\nS'inscrire_\n", font=("DIN", '14'
    ),bg='#ffffff').pack()

loginRegisterLabel=Label(self.fenetre2, text="Identifiant:",
    font=("DIN", '14'),bg='#ffffff')
loginRegisterLabel.pack()
loginRegisterForm=Entry(self.fenetre2, textvariable=self.
    loginR, width=30)
loginRegisterForm.pack()

passwordRegisterLabel=Label(self.fenetre2, text="\nMot_de_
    passe:", font=("DIN", '14'),bg='#ffffff')
passwordRegisterLabel.pack()
passwordRegisterForm=Entry(self.fenetre2, textvariable=self.
    passwordR, width=30)
passwordRegisterForm.pack()

bouton_envoyer=Button(self.fenetre2, text='S\'inscrire',
    command=self.registerAuth, fg='#000000')
bouton_envoyer.pack()

def main(self):
    self.fenetre1=Tk()
    self.fenetre1.title("MIG_SE_2013")
    self.fenetre1.attributes('-zoomed', 1)
    self.fenetre1.configure(background='white')
    titre=Label(self.fenetre1, text="\nMIG_SE_2013",font =("DIN", "34"
        ,"bold"), fg='#006eb8',bg='#ffffff')

```

```

titre.pack()
titre_logiciel=Label(self.fenetre1 , text="Reconnaissance_vocale\n"
                    ,font =("DIN" , "22"),bg='#ffffff ')
titre_logiciel.pack()

if not self.auth.connected:
    self.bouton_loginpopup=Button(self.fenetre1 ,text="Se_connecter
    _/_S'inscrire" , command=self.displayLogIn)
    self.bouton_loginpopup.pack()
    self.bouton_registerIn=Button(self.fenetre1 ,text="Enregistrer"
    , command=self.displayRecorder)
else:
    self.bouton_loginpopup=Button(self.fenetre1 ,text="Se_
    dÃ©connecter" , command=self.displayLogIn)
    self.bouton_loginpopup.pack()
    self.bouton_registerIn=Button(self.fenetre1 ,text="Enregistrer"
    , command=self.displayRecorder)
    self.displayRecorder()
    Button(self.fenetre1 ,text='Liste_des_mots_enregistrÃ©s' ,
    command=self.ouverture).pack()

bouton_fermer1=Button(self.fenetre1 ,text='Quitter' , command=self.
    fenetre1.destroy)
bouton_fermer1.pack()

self.recorderlabel=Label(self.fenetre1 ,bg='#ffffff ')
self.recorderlabel.pack()
espace3=Label(self.fenetre1 , text= '_\n_\n_' ,bg='#ffffff ')
espace3.pack()
photo=PhotoImage( file="speechapp/img/logomigSE.gif")      #insertion
    de l'image
labl = Label(self.fenetre1 , image=photo ,bg='#ffffff ')
labl.pack()
self.fenetre1.mainloop()
self.fenetre3.mainloop()

gui = Gui()
gui.main()

```

B. handling

B.1 fenetrehann.py

```
def hann_window_bis (signal):
    k = 0
    l = len(signal)
    j = 0
    liste = []
    while (k < l/(ecart_fenetre*RATE) and j < ((2*l)-(ecart_fenetre*RATE))):
        L = []
        for i in range(int(temps_fenetre*RATE)+1):
            L.append(signal[k*int(ecart_fenetre*RATE)+i]* \
0.5 * (1 - np.cos(2 * (np.pi) * (float(i) / (RATE * \
temps_fenetre)))))
            j += 1
        liste.append(L)
        k += 1
    return liste

def hann_window(signal):
    l = len(signal)
    k = 0
    liste = []
    while (k < l / (ecart_fenetre * RATE)):
        L = []
        for i in range(int(temps_fenetre * RATE) + 1):
            try:
                L.append(signal[k * int(ecart_fenetre * \
RATE) + i] * 0.5 * \
*(1 - np.cos(2 * np.pi * float(i) / (RATE * temps_fenetre) \
)))
            except IndexError:
                k = l/(ecart_fenetre*RATE)
                break
        liste.append(L)
        k+=1
    return liste

if __name__ == "__main__":
    #test de verification
    z = hann_window([250.*i/100000 for i in range (88200)])
    for i in range(len(z)):
        print(len(z[i]))
```

B.2 inverseDCT.py

```
TAILLE_TABLEAU_MEL_ENTREE = 24

NOMBRE_COMPOSANTES_GARDEES = 13

B = TAILLE_TABLEAU_MEL_ENTREE

def inverseDCTI(x): # x represente le tableau en mel donne par les
fonctions precedentes
    X = np.zeros(B)
    for k in range(B):
        X[k] = (0.5*(x[0]+math.pow(-1, k)*x[B-1]) + reduce(add, [x
            [n]*math.cos(math.pi*n*k/(B-1)) for n in range(1,B-1)
        ]))*math.sqrt(2./(B-1))
    return X

def inverseDCTII(x):
    X = [0 for i in range(B)]
    X[0] = reduce(add, [x[n] for n in range(B)]) / math.sqrt(B)
    for k in range(1, B):
        X[k] = reduce(add, [x[n]*math.cos(math.pi*(n+0.5)*k/B) for
            n in range(B)]) * math.sqrt(2./B)
    return X

def inverseDCTIII(x):
    X = np.zeros(B)
    for k in range(B):
        X[k] = (0.5*x[0]+reduce(add, [x[n]*math.cos(math.pi*n*(k
            +0.5)/B) for n in range(1,B)]))*math.sqrt(2./B)
    return X

if __name__ == "__main__":
    a = [math.cos(i) for i in range(24)]
    print(inverseDCTI(a))
    print(inverseDCTII(a))
    print(inverseDCTIII(a))
```

B.3 triangularFilterbank.py

```
def mel(f):
    return 2595*math.log(1+f/700.)/math.log(10)

def triangularFilter(tab, FE):
    """ Prend en paramètre une fenêtre de Hamming et la fréquence d'
    échantillonnage et retourne la fenêtre de Mel """
    pasOutput = mel(FE/2.)/12.
    pasFFT = FE/(2.*len(tab))
    outputTab = [0 for i in range(24)]
    for n in range(24):
        debut = n/2 * pasOutput + (n%2)*pasOutput/2.
```

```

    fin = n/2*pasOutput + (n%2)*pasOutput/2.+pasOutput
    milieu = (debut+fin)/2.
    for k in range(len(tab)):
        f = k*pasFFT
        if(mel(f) > debut and mel(f)<milieu):
            outputTab[n] += abs(((mel(f)-debut)/(
                pasOutput/2.))*tab[k])
        elif(mel(f) > milieu and mel(f) < fin):
            outputTab[n] += abs(((mel(f)- fin)/(
                pasOutput/2.) + 1)*tab[k])
        elif(mel(f) > fin):
            break
    for n in range(24):
        outputTab[n] = math.log(outputTab[n])
    return outputTab

```

B.4 passehaut.py

```

def passe_haut (X):
    N = len(X)
    Y = zeros(N, float)
    for i in range(1, N):
        Y[i] = X[i] - 0.95 * X[i - 1]
    Y[0] = X[0]
    return(Y)

```

B.5 fft.cpp

```

typedef std::complex<double> cDouble;

cDouble* listToTab(boost::python::list l)
{
    int N = boost::python::len(l);
    cDouble *t = (cDouble*) malloc(N*sizeof(cDouble));
    for (int i=0;i<N;i++)
        t[i] = boost::python::extract<cDouble>(l[i]);
    return t;
}

cDouble** listOfListToTab(boost::python::list l)
{
    int N = boost::python::len(l);
    cDouble** t = (cDouble**) malloc(N*sizeof(cDouble*));
    for (int i=0;i<N;i++)
        t[i] = listToTab(boost::python::extract<boost::python::list>(l[i]));
    return t;
}

bool is2Power(int N) { return N==1 || (N%2==0 && is2Power(N/2)); }

```



```

int get2Power(int N) { return pow(2, ceil(log(N)/log(2))); }

cDouble e(int k, int N) { return exp((cDouble)(-2j*M_PI*k/N)); }

boost::python::list tabToList(cDouble *t, int N)
{
    boost::python::list l = boost::python::list();
    for (int i=0;i<N;i++)
        l.append(t[i]);
    return l;
}

cDouble* fftCT(cDouble *sig)
{
    int i,j,k,p=0,f=1;
    int N = 1024;
    cDouble ekN;
    cDouble **tmp = (cDouble**) malloc(2*sizeof(cDouble*));

    for (i=0;i<2;i++)
        tmp[i] = (cDouble*) malloc(N*sizeof(cDouble));
    for (i=0;i<N;i++)
        tmp[0][i] = sig[i];
    for (i=N/2;i!=1;i/=2)
    {
        for (j=0;j<i;j++)
            for (k=0;k<N/(2*i);k++)
            {
                ekN = e(k,N/i)*tmp[p][i*(2*k+1)+j];
                tmp[f][i*k+j] = tmp[p][i*(2*k)+j] + ekN;
                tmp[f][i*k+j+N/2] = tmp[p][i*(2*k)+j] - ekN;
            }
        p = f;
        f = (p+1)%2;
    }

    free(tmp[f]);

    return tmp[p];
}

cDouble* fft(cDouble *sig, int N, int *sizeC, bool mid)
{
    cDouble* C;
    if (is2Power(N)) {
        C = fftCT(sig);
    } else {
        //std::cout<<"zPad needed";
        int NPadded = get2Power(N);
        cDouble* sigPadded = (cDouble*) malloc(NPadded*sizeof(cDouble));
        for (int i=0;i<N;i++)
            sigPadded[i] = sig[i];
        for (int i=N;i<NPadded;i++)

```

```

        sigPadded[i] = 0;
        C = fftCT(sigPadded);
        free(sigPadded);
    }

    if (mid) {
        int n = 512;
        cDouble *rep = (cDouble*) malloc(n*sizeof(cDouble));
        for (int i=0;i<n;i++)
            rep[i] = C[i];
        *sizeC = n;
        free(C);
        return rep;
    } else {
        *sizeC = N;
        return C;
    }
}

boost::python::list fftListe(boost::python::list pyEchs, bool mid=true)
{
    int nbEchs = boost::python::len(pyEchs);
    cDouble **echs = listOfListToTab(pyEchs);
    boost::python::list rep = boost::python::list();

    int sizeC;
    cDouble* C;

    for (int i=0;i<nbEchs-1;i++)
    {
        if (i%5==0) {
            //std::cout<<"Traitement du " << i << "eme
            //echantillon..." << std::endl;
        }
        C = fft(echs[i], 1024, &sizeC, mid);
        rep.append(tabToList(C, sizeC));
    }

    //std::cout << "Traitement du dernier echantillon..." << std::endl;
    int sizeLastEch = boost::python::len(pyEchs[nbEchs-1]);
    C = fft(echs[nbEchs-1], sizeLastEch, &sizeC, mid);
    rep.append(tabToList(C, sizeC));
    free(C);

    //std::cout << "Done !" << std::endl;
    return rep;
}

BOOST_PYTHON_MODULE(fft)
{
    using namespace boost::python;
    def("fftListe", fftListe);
}

```

C. HMM

C.1 creationVecteurHMM.py

```
D = TAILLE_FINALE_MFCC

#tabMel = [[(i+1)*(k+1) for i in range(D)] for k in range(3)] #liste des
#tableaux de mel de l'echantillon sonore

#print tabMel

def creeVecteur(tabMel, energyTable):
    choice = -1
    output = [[0 for i in range(D)] for k in range(len(tabMel))]
    for t in range(len(tabMel)):
        output[t] = tabMel[t][0:D]
        output[t][D-1] = energyTable[t]
    while( not choice in range(2) ):
        try:
            choice = 1
            #choice = int( input( "Voulez-vous incorporer les
            #differences premieres ? \n 0 : non    1 : oui :
            #" ))
        except NameError:
            print "Choix_non_valable"
    if (choice == 1):
        delta = [[0 for k in range(D)] for i in range(len(
            tabMel))]
        choice = -1
        for t in range(1, len(tabMel)):
            for k in range(D):
                delta[t][k] = output[t][k] - output[
                    t-1][k]
        for t in range(1, len(tabMel)):
            for k in range(D):
                output[t][k] += delta[t][k]
        #print "output :"
        #print output
        #print "delta :"
        #print delta
        while( not choice in range(2) ):
            try:
                choice = 1
                #choice = int( input( "Voulez-vous
                #incorporer les differences
```

```

                                secondes ? \n 0 : non 1 :
                                oui " ) )
except NameError:
    print "Choix_non_valable"
if (choice == 1):
    for t in range(1,len(tabMel)):
        for k in range(D):
            output[t][k] +=
                delta[t][k]-
                delta[t-1][k]

                                break
                                break
                                break
return output

```

C.2 markov.py

```

def getData(name):
    with open(name,"r") as f:
        content = pickle.Unpickler(f).load()

    return content

def writeData(name, data):
    with open(name, "w") as f:
        pickle.Pickler(f).dump(data)

def getID(d):
    l = []
    for i in range(d):
        l.append([])
        for j in range(d):
            if i==j:
                l[i].append(1.)
            else:
                l[i].append(0.)

    return l

def uniformPI(n):
    PI = []
    for i in range(n):
        PI.append(1./n)
    return PI

def uniformA(n):
    A = []
    for i in range(n):
        A.append(uniformPI(n))
    return A

def uniformC(n, m):
    C = []

```

```

    for i in range(n):
        C.append(uniformPI(m))
    return C

def uniformG_sigma(n, m, d):
    G_sigma = []
    for i in range(n):
        G_sigma.append([])
        for j in range(m):
            G_sigma[i].append(getID(d))
    return G_sigma

def normalize(l):
    t = 0
    for i in range(len(l)):
        t += l[i]*l[i]
    t = math.sqrt(t)

    if t == 0:
        return l

    for i in range(len(l)):
        l[i] /= t

    return l

def distL(a, b):
    r = 0
    for i in range(len(a)):
        r += (a[i]-b[i])*(a[i]-b[i])

    return math.sqrt(r)

def coupures(l):
    l_ = []
    for i in range(len(l)-1):
        l_.append(distL(l[i], l[i+1]))

    normalize(l_)

    coupures = []
    for i in range(len(l)-1):
        if l_[i] >= 0.35:
            coupures.append(i)

    return coupures

def add(l, l_):
    for i in range(len(l_)):
        done = False
        for j in range(len(l)):
            if l_[i] == l[j][0]:
                l[j] = (l[j][0], l[j][1]+1)

```

```

        done = True
        break
    elif l_[i] < l[j][0]:
        l.insert(j, (l_[i], 1))
        done = True
        break
    if done == False:
        l.append((l_[i], 1))
return l

def convert(l):
    max = 0
    for i in range(len(l)):
        if l[i][0] > max:
            max = l[i][0]

    l_ = [0 for k in range(max+1)]
    for i in range(len(l)):
        l_[l[i][0]] = l[i][1]

    return l_

def spikes(l):
    changed = True
    while changed == True:
        changed = False
        l_ = [0 for k in range(len(l))]

        if l[1] > l[0] and l[0] != 0:
            l_[1] += l[0]
            changed = True
        else:
            l_[0] += l[0]

        for i in range(1, len(l)-1):
            if l[i-1] > l[i] and l[i] != 0:
                l_[i-1] += l[i]
                changed = True
            elif l[i+1] > l[i] and l[i] != 0:
                l_[i+1] += l[i]
                changed = True
            else:
                l_[i] += l[i]

        if l[len(l)-2] > l[len(l)-1] and l[len(l)-1] != 0:
            l_[len(l)-2] += l[len(l)-1]
            changed = True
        else:
            l_[len(l)-1] += l[len(l)-1]

    l = l_

l_ = []

```

```

    for i in range(len(l)):
        if l[i] > 0:
            l_.append(i)

    return l_

def metaCoupures(seqs):
    l = []
    for i in range(len(seqs)):
        add(l, coupures(seqs[i]))

    l = convert(l)
    l = spikes(l)

    morceaux = [[] for k in range(len(l)+1)]
    for i in range(len(seqs)):
        k = 0
        for j in range(len(seqs[i])):
            morceaux[k].append(seqs[i][j])
            if k < len(l) and j == l[k]:
                k += 1

    mus = []
    for i in range(len(morceaux)):
        mu = [0 for k in range(len(morceaux[i][0]))]
        for j in range(len(morceaux[i])):
            for k in range(len(morceaux[i][j])):
                mu[k] += morceaux[i][j][k]

        for k in range(len(morceaux[i][0])):
            mu[k] /= len(morceaux[i])
        mus.append([mu])

    return mus

def buildHMMs(HMMs, HMMsPath, maxIt, path = "../db/hmm/"):
    hmm.clearHMMs()

    G_mu = []
    seqs = []
    for i in range(len(HMMsPath)):
        seqs.append([])
        G_mu.append([])
        for j in range(len(HMMsPath[i])):
            seqs[i].append(getData(path + HMMsPath[i][j]))
            G_mu[i] = G_mu[i] + metaCoupures(seqs[i][j]) # FAIL : this
                                                         # does not work for multi speaker : G_mu should contain
                                                         # different mus for each speaker
                                                         # FAIL : and the cutting of mus should be coherent

    d = 13

```

```

for i in range(len(HMMs)):
    n = len(G_mu[i])
    m = len(HMMsPath[i])
    hmm.createHMM(HMMs[i], HMMsPath[i], n, m, d, uniformPI(n),
        uniformA(n), uniformC(n, m), G_mu[i], uniformG_sigma(n, m, d))

    passSeqs = []
    for j in range(len(seqs[i])):
        passSeqs = passSeqs + seqs[i][j]
    x = hmm.baumWelch(HMMs[i], passSeqs, maxIt)
    if x == 0.5:
        print ("HMM '{}' _final_likelyhood_(log):_-inf".format(HMMs[i]))
        print ("WARNING_: _HMM_yielded_0_likelyhood_!")
    elif x == 0.8:
        print ("ERROR_: _HMM_'{}'_not_found!".format(HMMs[i]))
    elif x >= 1:
        print ("HMM '{}' _final_likelyhood_(log):_{}".format(HMMs[i],
            1-x))
        print ("WARNING_: _Baum-Welch_algorithm_ended_because_of_
            iterations '_limit_({})".format(maxIt))
    else:
        print ("HMM '{}' _final_likelyhood_(log):_{}".format(HMMs[i], x))
    print ("")

def loadHMMs(fileName):
    l = getData(fileName)
    hmm.setHMMs(l)

def saveHMMs(fileName):
    l = hmm.getHMMs()
    writeData(fileName, l)

def recognize(seq):
    l = hmm.recognize(seq)
    print ("Sequence_recognized_as:_{}_(log_probabilty:_{}_)".format(l[0],
        l[1]))
    return l[0]

def recognizeList(name, path):
    seqs = getData(path)
    for i in range(len(seqs)):

```

C.3 tableauEnergyPerFrame.py

```

def construitTableauEnergy(content):
    er = [0 for j in range(len(content))]
    for k in range(len(content)):
        er[k]=math.log(reduce(add,[abs(content[k][i])*abs(content[
            k][i]) for i in range(len(content))]))
    return er

```


C.4 hmm.cpp

```
const long double MIN_VALUE = 0.00001;

long double det(long double **sigma, int d) {
    long double r = 1;
    for (int i = 0; i < d; i++)
        r *= sigma[i][i];

    return r;
}

long double calcProduct(long double **sigma, long double *mu, long double*
x, int d) {
    long double r = 0;
    for (int i = 0; i < d; i++)
        r += (x[i]-mu[i])*(x[i]-mu[i])/sigma[i][i];

    return r;
}

void sumVects(long double ***seqs, long double ****gammas, int d, int sN,
int *sS, int i, int k, long double *r) {
    for (int a = 0; a < d; a++)
        r[a] = 0;

    for (int s = 0; s < sN; s++) {
        for (int t = 0; t < sS[s]; t++) {
            for (int a = 0; a < d; a++)
                r[a] += seqs[s][t][a]*gammas[s][t][i][k];
        }
    }
}

void sumMats(long double ***seqs, long double *mu, long double ****gammas,
int d, int sN, int *sS, int i, int k, long double **r) {
    for (int a = 0; a < d; a++) {
        for (int b = 0; b < d; b++)
            r[a][b] = 0;
    }

    for (int s = 0; s < sN; s++) {
        for (int t = 0; t < sS[s]; t++) {
            for (int a = 0; a < d; a++)
                r[a][a] += (seqs[s][t][a]-mu[a])*(seqs[s][t][a]-mu[a])*
                    gammas[s][t][i][k];
        }
    }
}

void mulVect(long double *v, long double a, int d, long double *r) {
    if (a == 0) { // Sum of sums is 0, so each sum is 0
        for (int i = 0; i < d; i++)
```

```

        r[i] = 200; // Far far far value so it's useless
    return;
}

    for (int i = 0; i < d; i++)
        r[i] = v[i]/a;
}

void mulMat(long double **m, long double a, int d, long double **r) {
    long double s = 0;
    for (int i = 0; i < d; i++)
        s += m[i][i]/a;

    for (int i = 0; i < d; i++) {
        for (int j = 0; j < d; j++) {
            if (i == j) {
                if (m[i][j] <= MIN_VALUE) // Cap min values
                    r[i][j] = MIN_VALUE;
                else
                    r[i][j] = m[i][j]/a;
            }
            else
                r[i][j] = 0;
        }
    }
}

class ContinuousHMM {
public:
    ContinuousHMM(std::string name, int n, int m, int d, std::vector<std::
        string> listSequences, long double *PI, long double**A, long
        double **C, long double ***G_mu, long double ****G_sigma);
    ~ContinuousHMM();

    void render();

    long double calcGaussianValue(long double **sigma, long double *mu,
        long double *x); // Calculates a single probability for a vector
        x_ in the gaussian sigma
    void calcProbabilitiesVector(long double *x, long double *r); //
        Calculate a probability for a vector x in each state's mixture
    void calcProbabilitiesSequence(long double **seq, int s, long double
        **prob); // Calculate probabilities for each vector of the
        sequence seq
    long double forward(long double **seq, int s, long double **prob, long
        double **alpha); // Implementation of the forward algorithm,
        returns the overall probability of the sequence
    void backward(long double **seq, int s, long double **prob, long
        double **beta); // Implementation of the backward algorithm (doesn
        't return overall probability)
    void calcXiOldGamma(long double **seq, int s, long double **alpha,
        long double **beta, long double p, long double **prob, long double
        ***xi, long double **oldGamma); // Calculates Xis and Old Gammas

```

```

    for latter calculus
void calcGamma(long double **seq, int s, long double **alpha, long
double **beta, long double **prob, long double ***gamma); //
    Calculates gamma for latter calculus
void calcSums(long double ***seqs, int sN, int *sS, long double ****
gamma, long double **littleSums, long double ***littleVect, long
double ****littleMat, long double *fatSums); // Calculates partial
sums for latter calculus
double baumWelch(long double ***seqs, int sN, int *sS, int maxIt =
100, int epsilon = 0.0000000001); // Baum-Welch Algorithm
    Implementation, learning algorithm

std::string name;
std::vector<std::string> listSequences;

int n;
int m;
int d;
long double *PI;
long double **A;
long double **C;
long double ***G_mu;
long double ****G_sigma;
};

ContinuousHMM::ContinuousHMM(std::string name, int n, int m, int d, std::
vector<std::string> listSequences, long double *PI, long double **A,
long double **C, long double ***G_mu, long double ****G_sigma) {
    this->name = name;
    this->n = n;
    this->m = m;
    this->d = d;
    this->listSequences = listSequences;
    this->PI = PI;
    this->A = A;
    this->C = C;
    this->G_mu = G_mu;
    this->G_sigma = G_sigma;
}

ContinuousHMM::~~ContinuousHMM() {
}

void ContinuousHMM::render() {
    std::cout << "Markov's Continuous Automaton" << name << std::endl;
    std::cout << "PI: ";
    for (int i = 0; i < n; i++) {
        if (i != n-1)
            std::cout << PI[i] << ", ";
        else
            std::cout << PI[i] << "]" << std::endl << std::endl;
    }
}

```

```

std::cout << "A_:" << std::endl;
for (int i = 0; i < n; i++) {
    std::cout << "[";
    for (int j = 0; j < n; j++) {
        if (j != n-1)
            std::cout << A[i][j] << ",";
        else
            std::cout << A[i][j] << "]" << std::endl;
    }
    if (i == n-1)
        std::cout << "]" << std::endl << std::endl;
}

std::cout << "C_:" << std::endl;
for (int i = 0; i < n; i++) {
    std::cout << "[";
    for (int j = 0; j < m; j++) {
        if (j != m-1)
            std::cout << C[i][j] << ",";
        else
            std::cout << C[i][j] << "]" << std::endl;
    }
    if (i == n-1)
        std::cout << "]" << std::endl << std::endl;
}

std::cout << "G_mu_:" << std::endl;
for (int i = 0; i < n; i++) {
    std::cout << "[";
    for (int j = 0; j < m; j++) {
        if (j != m-1) {
            std::cout << "[";
            for (int a = 0; a < d; a++)
                std::cout << G_mu[i][j][a] << ",";
            std::cout << "]" << std::endl;
        } else {
            std::cout << "[";
            for (int a = 0; a < d; a++)
                std::cout << G_mu[i][j][a] << ",";
            std::cout << "]" << std::endl;
            std::cout << "]" << std::endl;
        }
    }
    if (i == n-1)
        std::cout << "]" << std::endl << std::endl;
}

std::cout << "G_sigma_:" << std::endl;
for (int i = 0; i < n; i++) {
    std::cout << "[";
    for (int j = 0; j < m; j++) {
        if (j != m-1) {
            std::cout << "[";

```

```

        for (int a = 0; a < d; a++)
            std::cout << G_sigma[i][j][a][a] << ",_";
        std::cout << "]" << std::endl;
    } else {
        std::cout << "[";
        for (int a = 0; a < d; a++)
            std::cout << G_sigma[i][j][a][a] << ",_";
        std::cout << "]" << std::endl;
        std::cout << "]" << std::endl;
    }
}
if (i == n-1)
    std::cout << "]" << std::endl << std::endl;
}
}

long double ContinuousHMM::calcGaussianValue(long double **sigma, long
double *mu, long double *x) {
    long double den = sqrt(pow(2*M_PI, d) * det(sigma, d));
    long double num = exp((long double)-.5 * calcProduct(sigma, mu, x, d))
;

    if (num/den > 1) // Probability over 1, Markov's bullshit continuous
theory
        return 1.1;
    else
        return num/den;
}

void ContinuousHMM::calcProbabilitiesVector(long double *x, long double *r
) {
    for (int i = 0; i < n; i++) {
        r[i] = 0;
        for (int j = 0; j < m; j++)
            r[i] += C[i][j]*calcGaussianValue(G_sigma[i][j], G_mu[i][j], x
);
    }
}

void ContinuousHMM::calcProbabilitiesSequence(long double **seq, int s,
long double **prob) {
    for (int i = 0; i < s; i++)
        calcProbabilitiesVector(seq[i], prob[i]);
}

long double ContinuousHMM::forward(long double **seq, int s, long double
**prob, long double **alpha) {
    for (int i = 0; i < n; i++) // Setting for each state value at t=0
        alpha[0][i] = PI[i]*prob[0][i];

    for (int t = 1; t < s; t++) {
        for (int i = 0; i < n; i++) {
            long double r = 0;

```

```

        for (int j = 0; j < n; j++)
            r += alpha[t-1][j]*A[j][i];
        alpha[t][i] = r*prob[t][i];
    }
}

long double p = 0;
for (int i = 0; i < n; i++)
    p += alpha[s-1][i];

return p;
}

void ContinuousHMM::backward(long double **seq, int s, long double **prob,
    long double **beta) {
    for (int i = 0; i < n; i++) // Setting for each state value at t=s-1
        beta[s-1][i] = 1;

    for (int t = s-2; t >= 0; t--) {
        for (int i = 0; i < n; i++) {
            long double r = 0;
            for (int j = 0; j < n; j++)
                r += A[i][j]*beta[t+1][j]*prob[t+1][j];
            beta[t][i] = r;
        }
    }
}

// Uses xi and oldGamma. WARNING : xi and oldGamma must be defined !
void ContinuousHMM::calcXiOldGamma(long double **seq, int s, long double
    **alpha, long double **beta, long double p, long double **prob, long
    double ***xi, long double **oldGamma) {
    for (int t = 0; t < s-1; t++) {
        for (int i = 0; i < n; i++) {
            oldGamma[t][i] = 0;
            for (int j = 0; j < n; j++) {
                xi[t][i][j] = alpha[t][i]*A[i][j]*prob[t+1][j]*beta[t+1][j]
                    /p;
                oldGamma[t][i] += xi[t][i][j];
            }
        }
    }
}

void ContinuousHMM::calcGamma(long double **seq, int s, long double **
    alpha, long double **beta, long double **prob, long double ***gamma) {
    for (int t = 0; t < s; t++) {
        long double sumAB = 0;
        for (int i = 0; i < n; i++)
            sumAB += alpha[t][i]*beta[t][i];

        for (int i = 0; i < n; i++) {
            long double AB = alpha[t][i]*beta[t][i]/sumAB;

```

```

        for (int k = 0; k < m; k++) {
            if (prob[t][i] == 0) // This means the sum of
                                // probabilities is 0, thus a single one of them will be
                                // 0 too
                gamma[t][i][k] = 0;
            else
                gamma[t][i][k] = AB*C[i][k]*calcGaussianValue(G_sigma[
                    i][k], G_mu[i][k], seq[t])/prob[t][i];
        }
    }
}

// Uses littleSums, littleVect, littleMat, fatSums. WARNING : they must be
// defined !
void ContinuousHMM::calcSums(long double ***seqs, int sN, int *sS, long
    double ***gammas, long double **littleSums, long double ***littleVect
    , long double ***littleMat, long double *fatSums) {
    for (int i = 0; i < n; i++) {
        fatSums[i] = 0;
        for (int k = 0; k < m; k++) {
            littleSums[i][k] = 0;
            for (int a = 0; a < d; a++) {
                littleVect[i][k][a] = 0;
                for (int b = 0; b < d; b++)
                    littleMat[i][k][a][b] = 0;
            }

            for (int s = 0; s < sN; s++) {
                for (int t = 0; t < sS[s]; t++)
                    littleSums[i][k] += gammas[s][t][i][k];
            }
            sumVects(seqs, gammas, d, sN, sS, i, k, littleVect[i][k]);
            sumMats(seqs, G_mu[i][k], gammas, d, sN, sS, i, k, littleMat[i]
                [k]);
            fatSums[i] += littleSums[i][k];
        }
    }
}

double ContinuousHMM::baumWelch(long double ***seqs, int sN, int *sS, int
    maxIt, int epsilon) {
    long double oldLike = -1;
    long double like = 1;
    long double mean = 0;
    long double rap = 1;
    int it = 0;
    //bool decrease = false;
    int totalSize = 0;
    for (int s = 0; s < sN; s++)
        totalSize += sS[s];

    // Allocation of new model parameters

```

```

long double *_PI = (long double*) malloc(sizeof(long double)*n);
long double **_A = (long double**) malloc(sizeof(long double)*n*n);
long double **_C = (long double**) malloc(sizeof(long double)*n*m);
long double ***_G_mu = (long double***) malloc(sizeof(long double)*n*m*
d);
long double ****_G_sigma = (long double****) malloc(sizeof(long double)
*n*m*d*d);
for (int i = 0; i < n; i++) {
    _A[i] = (long double*) malloc(sizeof(long double)*n);
    _C[i] = (long double*) malloc(sizeof(long double)*m);
    _G_mu[i] = (long double**) malloc(sizeof(long double)*m*d);
    _G_sigma[i] = (long double****) malloc(sizeof(long double)*m*d*d);
    for (int j = 0; j < m; j++) {
        _G_mu[i][j] = (long double*) malloc(sizeof(long double)*d);
        _G_sigma[i][j] = (long double****) malloc(sizeof(long double)*d*d
);
        for (int a = 0; a < d; a++)
            _G_sigma[i][j][a] = (long double*) malloc(sizeof(long
double)*d);
    }
}

// Allocation of temporary arrays
long double ***alphas = (long double****) malloc(sizeof(long double)*
totalSize*n);
long double ***betas = (long double****) malloc(sizeof(long double)*
totalSize*n);
long double *ps = (long double*) malloc(sizeof(long double)*sN);
long double ***probs = (long double****) malloc(sizeof(long double)*
totalSize*n);

long double ****xis = (long double****) malloc(sizeof(long double)*
totalSize*n*n);
long double ***oldGammas = (long double****) malloc(sizeof(long double)*
totalSize*n);
long double ****gammas = (long double****) malloc(sizeof(long double)*
totalSize*n*n);

for (int s = 0; s < sN; s++) {
    alphas[s] = (long double****) malloc(sizeof(long double)*sS[s]*n);
    betas[s] = (long double****) malloc(sizeof(long double)*sS[s]*n);
    probs[s] = (long double****) malloc(sizeof(long double)*sS[s]*n);
    xis[s] = (long double****) malloc(sizeof(long double)*sS[s]*n*n);
    oldGammas[s] = (long double****) malloc(sizeof(long double)*sS[s]*n);
    gammas[s] = (long double****) malloc(sizeof(long double)*sS[s]*n*n);
    for (int t = 0; t < sS[s]; t++) {
        alphas[s][t] = (long double****) malloc(sizeof(long double)*n);
        betas[s][t] = (long double****) malloc(sizeof(long double)*n);
        probs[s][t] = (long double****) malloc(sizeof(long double)*n);
        xis[s][t] = (long double****) malloc(sizeof(long double)*n*n);
        oldGammas[s][t] = (long double****) malloc(sizeof(long double)*n);
        gammas[s][t] = (long double****) malloc(sizeof(long double)*n*n);
        for (int i = 0; i < n; i++) {

```



```

        xis[s][t][i] = (long double*) malloc(sizeof(long double)*n)
        ;
        gammas[s][t][i] = (long double*) malloc(sizeof(long double)
            *n);
    }
}

// Allocation of partial sums
long double **littleSums = (long double**) malloc(sizeof(long double)*n
    *m);
long double ***littleVect = (long double***) malloc(sizeof(long double)
    *n*m*d);
long double ****littleMat = (long double****) malloc(sizeof(long double)
    *n*m*d*d);
long double *fatSums = (long double*) malloc(sizeof(long double)*n);
for (int i = 0; i < n; i++) {
    littleSums[i] = (long double*) malloc(sizeof(long double)*m);
    littleVect[i] = (long double**) malloc(sizeof(long double)*m*d);
    littleMat[i] = (long double***) malloc(sizeof(long double)*m*d*d);
    for (int j = 0; j < m; j++) {
        littleVect[i][j] = (long double*) malloc(sizeof(long double)*d)
            ;
        littleMat[i][j] = (long double**) malloc(sizeof(long double)*d*
            d);
        for (int a = 0; a < d; a++)
            littleMat[i][j][a] = (long double*) malloc(sizeof(long
                double)*d);
    }
}

long double r = 0;
long double num = 0;
long double den = 0;
while (it < maxIt) {
    for (int s = 0; s < sN; s++)
        calcProbabilitiesSequence(seqs[s], sS[s], probs[s]);

    for (int s = 0; s < sN; s++) {
        ps[s] = forward(seqs[s], sS[s], probs[s], alphas[s]);
        backward(seqs[s], sS[s], probs[s], betas[s]);

        calcXiOldGamma(seqs[s], sS[s], alphas[s], betas[s], ps[s],
            probs[s], xis[s], oldGammas[s]);
        calcGamma(seqs[s], sS[s], alphas[s], betas[s], probs[s],
            gammas[s]);
    }

    // Calculation of sums
    calcSums(seqs, sN, sS, gammas, littleSums, littleVect, littleMat,
        fatSums);

    for (int i = 0; i < n; i++) {

```

```

// Setting PI and A
r = 0;
den = 0;
for (int s = 0; s < sN; s++) {
    r += oldGammas[s][0][i];
    for (int t = 0; t < sS[s]-1; t++)
        den += oldGammas[s][t][i];
}
_Pi[i] = r/sN;

if (den == 0) { // This means the probability to be in state i
    at time t is 0
    // So probabilities to go from i at t to j at t+1 will be
    0 for each j
    for (int j = 0; j < n; j++)
        _A[i][j] = (long double)(1./n); // Need to put equi-
        probabilities
}
else {
    for (int j = 0; j < n; j++) {
        num = 0;
        for (int s = 0; s < sN; s++) {
            for (int t = 0; t < sS[s]-1; t++)
                num += xis[s][t][i][j];
        }
        _A[i][j] = num/den;
    }
}

if (fatSums[i] == 0) { // C will be problematic
    for (int k = 0; k < m; k++)
        _C[i][k] = (long double)(1./m); // Need to put equi-
        probabilities
} else { // Setting C normally
    for (int k = 0; k < m; k++)
        _C[i][k] = littleSums[i][k]/fatSums[i];
}

// Setting G_mu and G_sigma
for (int k = 0; k < m; k++) {
    mulVect(littleVect[i][k], littleSums[i][k], d, _G_mu[i][k]
    );
    mulMat(littleMat[i][k], littleSums[i][k], d, _G_sigma[i][k]
    );
}
}

like = 1;
mean = 0;
for (int s = 0; s < sN; s++) {
    like *= ps[s];
    mean += ps[s];
}

```

```

mean /= sN;
//std::cout << "Likelyhood : " << like << " (" << mean << ")" <<
    std::endl << std::endl;
rap = 1;
if (oldLike != -1) {
    rap = like/oldLike;
    if (rap < 1) {
        //decrease = true;
        break;
    }
    else if (rap < (1+epsilon))
        break;
}

oldLike = like;

for (int i = 0; i < n; i++) {
    PI[i] = _PI[i];
    for (int j = 0; j < n; j++)
        A[i][j] = _A[i][j];
    for (int k = 0; k < m; k++) {
        C[i][k] = _C[i][k];

        for (int a = 0; a < d; a++) {
            G_mu[i][k][a] = _G_mu[i][k][a];
            G_sigma[i][k][a][a] = _G_sigma[i][k][a][a];
        }
    }
}

it++;
}

double result = 0;
if (oldLike == 0)
    result = 0.5;
else
    result = log10l(oldLike);

if (it == maxIt) {
    //std::cout << "Ended on max iteration" << std::endl;
    result = 1 - result;
}/* else {
    if (decrease)
        std::cout << "Ended on decreasing likelyhood" << std::endl;
    else
        std::cout << "Ended on stationary likelyhod" << std::endl;
}*/

//std::cout << "HMM '" << name << "' -> final likelyhood (iteration "
    << it << " of " << maxIt << ") : " << oldLike << " and mean : " <<
    mean << std::endl << std::endl;

```

```

// Freeings
// Freeing of new model parameters
for (int i = 0; i < n; i++) {
    for (int j = 0; j < m; j++) {
        for (int a = 0; a < d; a++)
            free(_G_sigma[i][j][a]);
        free(_G_mu[i][j]);
        free(_G_sigma[i][j]);
    }
    free(_A[i]);
    free(_C[i]);
    free(_G_mu[i]);
    free(_G_sigma[i]);
}
free(_PI);
free(_A);
free(_C);
free(_G_mu);
free(_G_sigma);

// Freeing of temporary arrays
for (int s = 0; s < sN; s++) {
    for (int t = 0; t < sS[s]; t++) {
        for (int i = 0; i < n; i++) {
            free(xis[s][t][i]);
            free(gammas[s][t][i]);
        }
        free(alphas[s][t]);
        free(betas[s][t]);
        free(probs[s][t]);
        free(xis[s][t]);
        free(oldGammas[s][t]);
        free(gammas[s][t]);
    }
    free(alphas[s]);
    free(betas[s]);
    free(probs[s]);
    free(xis[s]);
    free(oldGammas[s]);
    free(gammas[s]);
}
free(alphas);
free(betas);
free(ps);
free(probs);
free(xis);
free(oldGammas);
free(gammas);

// Freeing of partial sums
for (int i = 0; i < n; i++) {
    for (int j = 0; j < m; j++) {

```

```

        for (int a = 0; a < d; a++)
            free(littleMat[i][j][a]);
        free(littleVect[i][j]);
        free(littleMat[i][j]);
    }
    free(littleSums[i]);
    free(littleVect[i]);
    free(littleMat[i]);
}
free(littleSums);
free(littleVect);
free(littleMat);
free(fatSums);

return result;
}

std::vector<ContinuousHMM*> HMMs;

int findHMM(std::string s) {
    for (unsigned int i = 0; i < HMMs.size(); i++) {
        if (HMMs.at(i)->name.compare(s) == 0)
            return (int)i;
    }

    std::cout << "HMM_" << s << "_not_found_" << std::endl;
    return -1;
}

boost::python::list tabToList(long double *tab, int size) {
    boost::python::list _list;
    for (int i = 0; i < size; i++)
        _list.append(tab[i]);

    return _list;
}

boost::python::list tTabToLList(long double **tab, int size, int subSize)
{
    boost::python::list _list;
    for (int i = 0; i < size; i++)
        _list.append(tabToList(tab[i], subSize));

    return _list;
}

boost::python::list tTTabToLLList(long double ***tab, int size, int
subSize, int subSubSize) {
    boost::python::list _list;
    for (int i = 0; i < size; i++)
        _list.append(tTabToLList(tab[i], subSize, subSubSize));

    return _list;
}

```

```

}

boost::python::list tTTTabToLLLLList(long double *****tab, int size, int
    subSize, int subSubSize, int subSubSubSize) {
    boost::python::list _list;
    for (int i = 0; i < size; i++)
        _list.append(tTTTabToLLLLList(tab[i], subSize, subSubSize,
            subSubSubSize));

    return _list;
}

void createHMM(boost::python::str _name, boost::python::list
    _listSequences, int n, int m, int d, boost::python::list _PI, boost::
    python::list _A, boost::python::list _C, boost::python::list _G_mu,
    boost::python::list _G_sigma) {
    std::string name = boost::python::extract<std::string>(_name);
    long double *PI = (long double*)malloc(sizeof(long double)*n);
    long double **A = (long double**)malloc(sizeof(long double)*n*n);
    long double **C = (long double**)malloc(sizeof(long double)*n*m);
    long double ***G_mu = (long double***)malloc(sizeof(long double)*n*m*d
    );
    long double ****G_sigma = (long double****)malloc(sizeof(long double)*
    n*m*d*d);

    std::vector<std::string> listSequences;
    int numSeqs = boost::python::len(_listSequences);
    for (int i = 0; i < numSeqs; i++)
        listSequences.push_back(boost::python::extract<std::string>(
            _listSequences[i]));

    for(int i = 0; i < n; i++) {
        PI[i] = boost::python::extract<long double>(_PI[i]);

        A[i] = (long double*)malloc(sizeof(long double)*n);
        for (int j = 0; j < n; j++)
            A[i][j] = boost::python::extract<long double>(_A[i][j]);

        C[i] = (long double*)malloc(sizeof(long double)*m);
        G_mu[i] = (long double**)malloc(sizeof(long double)*m*d);
        G_sigma[i] = (long double***)malloc(sizeof(long double)*m*d*d);
        for (int j = 0; j < m; j++) {

            C[i][j] = boost::python::extract<long double>(_C[i][j]);

            G_mu[i][j] = (long double*)malloc(sizeof(long double)*d);
            G_sigma[i][j] = (long double***)malloc(sizeof(long double)*d*d
            );
            for (int k = 0; k < d; k++) {
                G_mu[i][j][k] = boost::python::extract<long double>(_G_mu[
                    i][j][k]);
            }
        }
    }
}

```

```

        G_sigma[i][j][k] = (long double*) malloc(sizeof(long double)
        *d);
        for (int l = 0; l < d; l++)
            G_sigma[i][j][k][l] = boost::python::extract<long
            double>(_G_sigma[i][j][k][l]);
    }
}

ContinuousHMM *M = new ContinuousHMM(name, n, m, d, listSequences, PI,
    A, C, G_mu, G_sigma);
HMMs.push_back(M);
}

void createHMMFromList(boost::python::list _HMM) {
    boost::python::str _name = boost::python::extract<boost::python::str>(_HMM[0]);
    boost::python::list _listSequences = boost::python::extract<boost::python::list>(_HMM[1]);
    int n = boost::python::extract<int>(_HMM[2]);
    int m = boost::python::extract<int>(_HMM[3]);
    int d = boost::python::extract<int>(_HMM[4]);
    boost::python::list _PI = boost::python::extract<boost::python::list>(_HMM[5]);
    boost::python::list _A = boost::python::extract<boost::python::list>(_HMM[6]);
    boost::python::list _C = boost::python::extract<boost::python::list>(_HMM[7]);
    boost::python::list _G_mu = boost::python::extract<boost::python::list>(_HMM[8]);
    boost::python::list _G_sigma = boost::python::extract<boost::python::list>(_HMM[9]);

    createHMM(_name, _listSequences, n, m, d, _PI, _A, _C, _G_mu, _G_sigma);
}

void setHMMs(boost::python::list _HMMs) {
    int size = boost::python::len(_HMMs);
    for (int i = 0; i < size; i++) {
        createHMMFromList(boost::python::extract<boost::python::list>(_HMMs[i]));
    }
}

boost::python::list getHMMs() {
    boost::python::list _HMMs;
    for (unsigned int i = 0; i < HMMs.size(); i++) {
        boost::python::list _HMM;
        _HMM.append(boost::python::str(HMMs.at(i)->name));
        boost::python::list _listSequences;
        for (unsigned int j = 0; j < HMMs.at(i)->listSequences.size(); j
            ++)
```

```

        _listSequences.append(HMMs.at(i)->listSequences.at(j))
        ;
        _HMM.append(_listSequences);
        int n = HMMs.at(i)->n;
        int m = HMMs.at(i)->m;
        int d = HMMs.at(i)->d;
        _HMM.append(n);
        _HMM.append(m);
        _HMM.append(d);
        _HMM.append(tabToList(HMMs.at(i)->PI, n));
        _HMM.append(tTabToLLList(HMMs.at(i)->A, n, n));
        _HMM.append(tTabToLLList(HMMs.at(i)->C, n, m));
        _HMM.append(tTTabToLLList(HMMs.at(i)->G_mu, n, m, d));
        _HMM.append(tTTabToLLList(HMMs.at(i)->G_sigma, n, m, d, d));

        _HMMs.append(_HMM);
    }

    return _HMMs;
}

void clearHMMs() {
    while (HMMs.size() != 0) {
        ContinuousHMM *M = HMMs.at(0);
        HMMs.erase(HMMs.begin());
        delete M;
    }
}

void removeHMM(boost::python::str _name) {
    std::string name = boost::python::extract<std::string>(_name);
    int id = findHMM(name);
    if (id == -1)
        return;

    ContinuousHMM *M = HMMs.at(id);
    HMMs.erase(HMMs.begin()+id);
    delete M;
}

void renderHMM(boost::python::str _name) {
    std::string name = boost::python::extract<std::string>(_name);
    int id = findHMM(name);
    if (id != -1)
        HMMs.at(id)->render();
}

long double forward(boost::python::str _name, boost::python::list _seq) {
    std::string name = boost::python::extract<std::string>(_name);
    int id = findHMM(name);
    if (id == -1)
        return -1;

```



```

int s = boost::python::len(_seq);

long double **seq = (long double**) malloc(sizeof(long double)*s*HMMs.
    at(id)->d);
for (int t = 0; t < s; t++) {
    seq[t] = (long double*) malloc(sizeof(long double)*HMMs.at(id)->d);
    for (int i = 0; i < HMMs.at(id)->d; i++)
        seq[t][i] = boost::python::extract<long double>(_seq[t][i]);
}

long double p;
long double **prob = (long double**) malloc(sizeof(long double)*s*HMMs.
    at(id)->n);
long double **alpha = (long double**) malloc(sizeof(long double)*s*HMMs.
    .at(id)->n);
for (int t = 0; t < s; t++) {
    prob[t] = (long double*) malloc(sizeof(long double)*HMMs.at(id)->n)
        ;
    alpha[t] = (long double*) malloc(sizeof(long double)*HMMs.at(id)->n
        );
}

HMMs.at(id)->calcProbabilitiesSequence(seq, s, prob);
p = HMMs.at(id)->forward(seq, s, prob, alpha);

for (int t = 0; t < s; t++) {
    free(seq[t]);
    free(prob[t]);
    free(alpha[t]);
}
free(seq);
free(prob);
free(alpha);

std::cout << "Forward_:_" << p << std::endl;
return p;
}

double baumWelch(boost::python::str _name, boost::python::list _seqs, int
it) {
    std::string name = boost::python::extract<std::string>(_name);
    int id = findHMM(name);
    if (id == -1)
        return 0.8;

    int sN = boost::python::len(_seqs);

    int *sS = (int*) malloc(sizeof(int)*sN);
    int totalSize = 0;
    for (int s = 0; s < sN; s++) {
        sS[s] = boost::python::len(boost::python::extract<boost::python::
            list>(_seqs[s]));
        totalSize += sS[s];
    }
}

```

```

}

long double ***seqs = (long double**) malloc(sizeof(long double)*
    totalSize*HMMs.at(id)->d);
for (int s = 0; s < sN; s++) {
    seqs[s] = (long double**) malloc(sizeof(long double)*sS[s]*HMMs.at(
        id)->d);
    for (int t = 0; t < sS[s]; t++) {
        seqs[s][t] = (long double*) malloc(sizeof(long double)*HMMs.at(
            id)->d);
        for (int i = 0; i < HMMs.at(id)->d; i++)
            seqs[s][t][i] = boost::python::extract<long double>(_seqs[
                s][t][i]);
    }
}

double d = HMMs.at(id)->baumWelch(seqs, sN, sS, it);

for (int s = 0; s < sN; s++) {
    for (int t = 0; t < sS[s]; t++) {
        free(seqs[s][t]);
    }
    free(seqs[s]);
}
free(seqs);
free(sS);

return d;
}

boost::python::list recognize(boost::python::list _seq) {
    int s = boost::python::len(_seq);

    int d = 13;
    long double **seq = (long double**) malloc(sizeof(long double)*s*d);
    for (int t = 0; t < s; t++) {
        seq[t] = (long double*) malloc(sizeof(long double)*d);
        for (int i = 0; i < d; i++)
            seq[t][i] = boost::python::extract<long double>(_seq[t][i]);
    }

    long double p;
    long double maxP = 0;
    std::string maxName;
    long double **prob;
    long double **alpha;

    for (int id = 0; id < (int)HMMs.size(); id++) {
        prob = (long double**) malloc(sizeof(long double)*s*HMMs.at(id)->n);
        ;
        alpha = (long double**) malloc(sizeof(long double)*s*HMMs.at(id)->n);
        for (int t = 0; t < s; t++) {

```

```

        prob[t] = (long double*) malloc(sizeof(long double)*HMMs.at(id)
->n);
        alpha[t] = (long double*) malloc(sizeof(long double)*HMMs.at(id)
->n);
    }

    HMMs.at(id)->calcProbabilitiesSequence(seq, s, prob);
    p = HMMs.at(id)->forward(seq, s, prob, alpha);

    if (p > maxP) {
        maxP = p;
        maxName = HMMs.at(id)->name;
    }

    for (int t = 0; t < s; t++) {
        free(prob[t]);
        free(alpha[t]);
    }
    free(prob);
    free(alpha);
}

for (int t = 0; t < s; t++)
    free(seq[t]);
free(seq);

boost::python::list result;
result.append(maxName);
result.append(log10(maxP));

return result;
}

BOOST_PYTHON_MODULE(hmm)
{
    using namespace boost::python;
    def("createHMM", createHMM);
    def("setHMMs", setHMMs);
    def("getHMMs", getHMMs);
    def("clearHMMs", clearHMMs);
    def("removeHMM", removeHMM);
    def("renderHMM", renderHMM);
    def("forward", forward);
    def("baumWelch", baumWelch);
    def("recognize", recognize);
}

```

D. recorder

D.1 recorder.py

```
def recorder(db, dirName="", nbRecording=-1, askForWord=True, seconds=-1,
            nbWords=1, fileName="", confirm=True):
    """ ProcÃde Ã l'enregistrement """

    if nbRecording < 0:
        nbRecording = raw_input("Combien d'enregistrement par mots
                                _?_")
        nbRecording = int(nbRecording)
    n = ""
    for k in range(nbWords):
        if askForWord:
            mot = raw_input("Entrez le mot Ã enregistrer:_:_")
        else:
            mot = ""

        if seconds < 0:
            seconds = raw_input("Entrez le nombre de secondes
                                pour l'enregistrement:_:_")
            seconds = float(seconds) + 1

        for i in range(nbRecording):
            if confirm:
                raw_input("Appuyez sur une touche pour
                            commencer l'enregistrement:_:_")
            p = pyaudio.PyAudio()

            stream = p.open(format=FORMAT,
                            channels=CHANNELS,
                            rate=RATE,
                            input=True,
                            frames_per_buffer=CHUNK)

            #print "Enregistrement[" , i , "]"

            frames = []

            for j in range(0, int(RATE / CHUNK * seconds)):
                data = stream.read(CHUNK)
                frames.append(data)

            #print "Fin - Enregistrement[" , i , "]"
```

```

        stream.stop_stream()
        stream.close()
        p.terminate()
        if dirName != "":
            random.seed()
            if fileName == "":
                fileName = hashlib.sha224(str(
                    random.randint(0,1e10))).
                   hexdigest()
            name = dirName + "/" + fileName + ".wav"
        else:
            n = str(i)
            name = mot + "/" + n + ".wav"
        db.addWave(name, CHANNELS, p.get_sample_size(FORMAT)
            , RATE, frames, p)
        #print "Fin du mot ", i

    return fileName

```

D.2 sync.py

```

def sync(amplitudes):
    tOut = 400
    N = len(amplitudes)
    coeff_lissage = 5
    max = 0
    for i in range(N):
        if abs(amplitudes[i]) > max:
            max = abs(amplitudes[i])

    #print("Max is {}".format(max))

    seuilFor = max/8
    seuilBack = max/7
    #print("Seuil is {}".format(valeurSeuil))

    maxDiff = 300
    maxRemove = 800
    #print("MaxDiff is {}".format(maxDiff))

    iMin = -1
    iMin2 = -1
    iMax = -1
    iMax2 = -1
    inIt = False
    lastHit = -1

    for i in range(N):
        if iMin == -1 and amplitudes[i] > seuilFor:
            iMin = i
            lastHit = i
            inIt = True

```

```

    if iMin != -1:
        if i - iMin > maxRemove: # Won't remove more than maxRemove
            break
        elif inIt == True and amplitudes[i] > seuilFor:
            lastHit = i
        elif inIt == True and amplitudes[i] < seuilFor and i - lastHit
            >= maxDiff:
            inIt = False
        elif inIt == False and amplitudes[i] > seuilFor:
            iMin2 = i
            break

inIt = False
lastHit = -1

for i in range(N-1, -1, -1):
    if iMax == -1 and amplitudes[i] > seuilBack:
        iMax = i
        lastHit = i
        inIt = True
    if iMax != -1:
        if iMax - i > maxRemove: # Won't remove more than maxRemove
            break
        elif inIt == True and amplitudes[i] > seuilBack:
            lastHit = i
        elif inIt == True and amplitudes[i] < seuilBack and lastHit -
            i >= maxDiff:
            inIt = False
        elif inIt == False and amplitudes[i] > seuilBack:
            iMax2 = i
            break

#print("iMin is {}".format(iMin))
#print("iMin2 is {}".format(iMin2))
#print("iMax is {}".format(iMax))
#print("iMax2 is {}".format(iMax2))
#print("")

if iMin2 != -1 and iMin2-iMin > tOut:
    iMin2 -= tOut
if iMax2 != -1 and iMax-iMax2 > tOut:
    iMax2 += tOut

if iMin > tOut:
    iMin -= tOut
if N-1 - iMax > tOut:
    iMax += tOut

amplitudes_coupe = [0. for i in range(iMax-iMin+1)]
for i in range(iMax-iMin+1):
    amplitudes_coupe[i] = amplitudes[iMin + i]

```

```

if iMin2 != -1 or iMax2 != - 1:
    if iMin2 == -1:
        iMin2 = iMin
    if iMax2 == -1:
        iMax2 = iMax

    if iMin2 - iMin <= maxRemove and iMax - iMax2 <= maxRemove:
        return amplitudes_coupe
    amplitudes_coupe2 = [0. for i in range(iMax2-iMin2+1)]
    for i in range(iMax2-iMin2+1):
        amplitudes_coupe2[i] = amplitudes[iMin2 + i]

    return sync(amplitudes_coupe2)
else:
    return amplitudes_coupe
def syncFile(path, name, prefix = "sync_"):
    #print("Synching : {}".format(name))
    ampli = scipy.io.wavfile.read(path + name)
    ampli2 = sync(ampli[1])
    scipy.io.wavfile.write(path + prefix + name, ampli[0], int16(ampli2))
    #print("Done\n\n")

def cutBeginning(path, name, prefix = "cut_"):
    ampli = scipy.io.wavfile.read(path + name)
    ampli2 = ampli[1][22050:]
    scipy.io.wavfile.write(path + prefix + name, ampli[0], int16(ampli2))

def sox_handling(fileName, noiseName, pathToTmp = "../db/waves/tmp/"):
    pass
    #os.system('sox "' + noiseName + '" -n noiseprof "' + pathToTmp + 'noise.prof'')
    #os.system('sox "' + fileName + '" "' + fileName + '" noisered "' + pathToTmp + 'noise.prof' 0.21')
    #os.remove(pathToTmp + "noise.prof")

if __name__ == "__main__":
    syncFile("3_0")
    syncFile("3_1")
    syncFile("5_0")
    syncFile("5_1")

```

E. utils

E.1 animate.py

```
data = Cs          # Liste de listes des valeurs
xMin = 350         # Echelles d'affichage
xMax = 550
yMin = 0
yMax = 100
interv = 50        # Millisecondes entre chaque image

fram = len(data)
fig = plt.figure()
ax = plt.axes(xlim=(xMin, xMax), ylim=(yMin, yMax))
line, = ax.plot([], [], lw=2)

def init():
    line.set_data([], [])
    return line,

def animate(i):
    L = len(data[i])
    x = np.linspace(0, L-1, L)
    y = [abs(data[i][int(k)]) for k in x]
    line.set_data(x, y)
    return line,

anim = animation.FuncAnimation(fig, animate, init_func=init,
                               frames=fram, interval=interv, blit=True)
```

E.2 constantes.py

```
# pour le recorder:
CHUNK = 1024 # nombre de bits enregistrÃ©s par boucle
FORMAT = pyaudio.paInt16
CHANNELS = 1 # On est en mono
RATE = 44100 #FrÃ©quence

# Synchro:
COEFF_LISSAGE = 5 # Ã dÃ©terminer empiriquement
T_MIN = 50 # blanc minimum avant le son
COEFF_COUPE = 0.0000001 # en pourcent
```



```
# fenetre hann :

ecart_fenetre = 0.01301587
temps_fenetre = 0.023219954648526

#Creation MFCC

TAILLE_FINALE_MFCC = 13

NB_ITERATIONS = 10
```

E.3 db.py

```
class Db:
    """ Files manager to store .wav and hmm
        Attributes :
            -> filesList : name of the file containing the list of files
                stored """

    filesListName = "filesList"
    prefixPath = ""
    verbose = False

    def __init__(self, prefixPath = "", filesListName = "filesList",
        verbose = False):
        """ Constructor which needs prefix of the directory which contains
            (or will contain) the stored files """
        Db.prefixPath = prefixPath
        Db.filesListName = filesListName
        Db.verbose = verbose
        self.log = ""
        try:
            with open(Db.prefixPath + Db.filesListName + ".txt","r") as f:
                self.filesList = pickle.Unpickler(f).load()
        except IOError:
            Db.reset(True)
            raise Exception("L'instanciation_a_ete_annulee_car_le_fichier_
                de_gestion_de_la_base_de_donnees_n'existe_pas")

    def getFile(self, fileName, dirFile=""):
        """ Add a file to the list of files handled par the database
            system
            Parameters :
                @fileName : name of the file in the storage directory
                    prefixed by dirFile
                @dirFile : add a prefix to files and give others handling
                    available
            """
        if len(dirFile) == 0:
            dirFile = "storage"
```

```

if fileName in self.filesList:
    try:
        if dirFile == "waves":
            content = scipy.io.wavfile.read(Db.prefixPath +
                dirFile + "/" + fileName)
            return content
        elif fileName in self.filesList:
            with open(Db.prefixPath + dirFile + "/" + fileName, "r"
                ) as f:
                content = pickle.Unpickler(f).load()
            return content
        except IOError:
            raise Exception("La_lecture_de_fichier_a_echoue")
    else:
        self.addLog("le_fichier_n'est_pas_gere_pas_la_base_de_donnees"
            )
        return ""

def getWaveFile(self, fileName):
    """ Alias of getFile for .wav """
    return self.getFile(fileName, "waves")

def addWave(self, fileName, CHANNELS, FORMAT, RATE, frames, p):
    if os.access(Db.prefixPath + "waves/" + fileName, os.F_OK):
        pass
        #Il faudrait rajouter la gestion de l'existence de deux mÃªmes
        fichiers
    dirName = os.path.dirname(fileName)
    if not os.access(Db.prefixPath + "waves/" + dirName, os.F_OK):
        os.mkdir(Db.prefixPath + "waves/" + dirName)
    wf = wave.open(Db.prefixPath + "waves/" + fileName, 'wb')
    wf.setnchannels(CHANNELS)
    wf.setsampwidth(FORMAT)
    wf.setframerate(RATE)
    wf.writeframes(b''.join(frames))
    wf.close()

    self.addFileToList(fileName, "waves")
    self.syncToFile()

def addWaveFromAmp(self, fileName, freq, amp, dirName="waves/", addToList=
    True):
    scipy.io.wavfile.write(Db.prefixPath + dirName + fileName, freq,
        int16(amp))
    if addToList:
        self.addFileToList(fileName, "waves")

```

```

        self.syncToFile()

def addFileToList(self, fileName, dirFile=""):
    """ Add a file to the list. Needs that the file already exists
        Parameters :
            @fileName : name of the file in the storage directory
                        prefixed by dirFile
            @dirFile : prefix of the file
    """
    if len(dirFile) == 0:
        dirFile = "storage"
    if os.access(Db.prefixPath + dirFile + "/" + fileName, os.F_OK):
        if not fileName in self.filesList:
            self.filesList.append(fileName)
            self.syncToFile()
            self.addLog("L'insertion_du_fichier_a_bien_ete_effectuee")
        else:
            self.addLog("Le_fichier_est_deja_dans_la_bibliothèque")
    else:
        self.addLog("Le_fichier_n'existe_pas")

def addFile(self, fileName, content, dirFile=""):
    """ Add a file to the list and to the storage directory
        Parameters :
            @fileName : name of the file in the storage directory
                        prefixed by dirFile
            @content : the content to pickle in the file
            @dirFile : prefix of the file
    """
    if len(dirFile) == 0:
        dirFile = "storage"
    with open(Db.prefixPath + dirFile + "/" + fileName, "w") as f:
        pickle.Pickler(f).dump(content)
    self.addFileToList(fileName)

def deleteFileFromList(self, fileName, dirFile=""):
    """ Remove a file from the list but does NOT remove the file from
        the disk
        Parameters :
            @fileName : name of the file in the storage directory
                        prefixed by dirFile
            @dirFile : prefix of the file
    """
    if len(dirFile) == 0:
        dirFile = "storage"
    if fileName in self.filesList:
        self.filesList.remove(fileName)
    try:
        dirName = os.path.dirname(fileName)

```

```

        if os.access(Db.prefixPath + dirFile + "/" + fileName, os.F_OK):
            os.remove(Db.prefixPath + dirFile + "/" + fileName)
            self.addLog("Le_fichier_a_bien_ete_supprime")
        else:
            self.addLog("Le_fichier_n'existe_pas")
    except OSError:
        self.addLog("La_suppression_a_echoue")
    try:
        os.rmdir(Db.prefixPath + dirFile + "/" + dirName)
        self.addLog("Le_dossier_a_bien_ete_supprime")
    except OSError:
        pass
    self.syncToFile()
    self.addLog("La_suppression_du_fichier_a_bien_ete_effectuee")
else:
    self.addLog("Le_fichier_n'existe_pas_ou_n'est_pas_gere_par_la_base_de_donnees")

def syncToFile(self):
    """ Synchronize the list of the files stored from the current attribute """
    try:
        with open(Db.prefixPath + Db.filesListName + ".txt", "w") as f:
            pickle.Pickler(f).dump(self.filesList)
    except IOError:
        raise Exception("Le_fichier_n'existe_pas")

def recursiveSync(self, dirName = "", dirIni = "storage/"):
    """ Synchronize the list of the files stored by studying recursively the current tree """
    for f in os.listdir(Db.prefixPath + dirIni + dirName):
        if os.path.isfile(os.path.join(Db.prefixPath + dirIni + dirName, f)):
            self.addFileToList(os.path.join(dirName, f), dirIni)
        else:
            self.sync(dirName + f + "/", dirIni)

def sync(self, dirName = "", dirIni = "storage/"):
    #Delete files that don't exist anymore in the list
    for k, f in enumerate(self.filesList):
        if not os.access(Db.prefixPath + dirIni + f, os.F_OK) and not os.access(Db.prefixPath + "storage/" + f, os.F_OK):
            del self.filesList[k]
    self.addFile(Db.filesListName + ".txt", self.filesList)
    self.recursiveSync(dirName, dirIni)

def reset(force=False):
    """ Reset the files list """

```

```

msg = "etes-vous_sur_de_vouloir_reinitialiser_la_liste_des_
      fichiers_?(Oui_=0/Non_=1)"
if force or int(input(msg)) == 0:
    with open(Db.prefixPath + Db.filesListName + ".txt","w") as f:
        c = pickle.Pickler(f)
        c.dump([])
        print "Reinitialisation_reussie_pour_les_fichiers"
reset = staticmethod(reset)

def syncHmm(self):
    hmmList = self.getFile("hmmList.txt")
    for k,f in hmmList.items():
        if not os.access(Db.prefixPath + "hmm/" + f,os.F_OK):
            del hmmList[k]
    self.addFile("hmmList.txt", hmmList)

def printFilesList(self,dirName="",printBool=True,*extRequired):
    """ Display the files list
        Parameters:
            @dirName : a prefixed
            @*extRequired : contains the extensions to display (e.g.
                           <.wav, .txt>) """
    filesListExt = []
    n = 0
    for k,f in enumerate(self.filesList):
        #On recupère l'extension du fichier parcouru
        a,ext = os.path.splitext(f)
        d = os.path.dirname(f)
        if (dirName == "" or d == dirName ) and (len(extRequired) == 0
            or ext in extRequired):
            if printBool:
                print n, "_", k, "_", f
                n += 1
            filesListExt.append(f)
    return filesListExt

def printDirFiles(self,dirName="storage/"):
    """ Display the list of files in a directory
        Parameters :
            @dirName = "storage/" : directory to browse """
    dirListExt = []
    l = os.listdir(Db.prefixPath + dirName)
    for k,f in enumerate(l):
        #On recupère l'extension du fichier parcouru
        print k, "_", f
    return l

def __str__(self):
    print self.filesList

def addLog(self,s,fileName=""):

```

```

        if Db.verbose:
            print s
            self.log += "\n" + s
            #self.addFile("dblog" + fileName + ".txt", self.log, "logs/")

    def logDump(self, fileName, log=""):
        if log == "":
            name = "dblog"
            log = self.log
        else:
            name = "handlinglog"
        self.addFile(name + fileName + ".txt", log, "logs/")

if __name__ == "__main__":
    db = Db()
    db.addFileToList("test.txt")
    db.getFile("test.txt")
    print(db)

```

E.4 util.py

```

def is2Power(N):
    return N == np.power(2, int(math.log(N, 2)))

def get2Power(N):
    return int(np.power(2, int(math.log(N, 2)) + 1))

def zPad(sig):
    N = len(sig)
    return sig + [0 for i in range(get2Power(N) - N)]

def reduc(M,N):
    d = gcd(M, N)
    return M/d, N/d

def pgcd(a,b):
    # une fonction fractions.gcd(a, b) est d'implémentée dans
    Python
    return gcd(a, b)

def W(k,N):
    return np.exp(-(2*np.pi*k/N)*1j)

def restreindre(sig):
    M = float(max(abs(sig)))
    return [float(sig[k])/M for k in range(len(sig))]

def getSin(freq, N, freqEch=44100):
    return [np.sin(2*np.pi*freq*t/freqEch) for t in range(N)]

```

F. SpeechApp

F.1 main

```
navigator.getUserMedia = (navigator.getUserMedia ||
                          navigator.webkitGetUserMedia ||
                          navigator.mozGetUserMedia);
window.AudioContext = window.AudioContext || window.webkitAudioContext;
window.URL = window.URL || window.webkitURL || window.mozURL;

var mediaRecorder; //Object MediaRecorder
//var audioElement = document.getElementById('audio'); //L'object audio
pour le direct play
var mediaStream; //Le flux LocalMediaStream pour moz browsers

var webkitaudio_context;
var webkitrecorder;

var recording = false;
var nav = null; //Enregistre le type de navigateur: moz ou webkit

var user = "demo";
var hashedPass = "8b1c1c1eae6c650485e77efbc336c5bfb84ffe0b0bea65610b721762";
var clientDB = "demo";
var SERVERURL = 'localhost:8010';

onload = function() {
    /* Au chargement, si l'API MediaRecorder est supportÃ©e,
on initialise l'entrÃ©e audio avec l'API getUserMedia */
    if (navigator.getUserMedia){
        if (typeof MediaRecorder === 'undefined'){
            if (navigator.getUserMedia && window.AudioContext && window.
                URL){
                nav = 'webkit';
                webkitaudio_context = new AudioContext;
            }
        }
        else{
            alert("Votre navigateur ne nous supporte pas :Ã°(");
        }
    }
}
```

```

    }
    else{
        nav = 'moz';
    }
}

if (nav !== null){
    navigator.getUserMedia({audio: true},
        initRecording,
        function(err) {
            console.log("The_following_error_occured:" +
                err);
        }
    );
    console.log(nav + '_compatibility_mode_running...');
}
};

function initRecording(localMediaStream){
    if (nav === 'moz'){
        mozinitRecording(localMediaStream);
    }
    else if (nav === 'webkit'){
        webkitinitRecording(localMediaStream);
    }
}

function main(){
    /* Decide quelle action lancer lorsque le bouton est toggÃ© */
    var microphone = document.getElementById('microphone');
    if (!recording){
        try{
            microphone.className = "wobble_animated";
            microphone.style.border = '5px_solid_#003173';
            startRecord();
            recording = true;
            //changeLogoBG('green');
        }
        catch (e){
            console.log("Recording_issue\n" + e);
        }
    }
    else{
        try{
            stopRecord();
            recording = false;
            //changeLogoBG('white');
            microphone.className = "";
            microphone.style.border = '5px_solid_white';
        }
    }
}

```



```

        catch (e){
            console.log("Recording_stop_issue\n" + e);
        }
    }

}

function startRecord(){
    /* Lance un enregistrement */
    navSwitch(mozstartRecorder, webkitstartRecorder);
    console.log('recording');
}

function stopRecord(){
    /* Stopper et clore un enregistrement */
    navSwitch(mozstopRecorder, webkitstopRecorder);
}

function navSwitch(mozaction, webkitaction){
    console.log(nav);
    if (nav == 'moz'){
        mozaction();
    }
    else if (nav == 'webkit'){
        webkitaction();
    }
}

function preInteract(audioBlob, blobType){
    if (nav == 'webkit'){
        var url = URL.createObjectURL(audioBlob);
    }
    else if (nav == 'moz'){
        var url = window.URL.createObjectURL(audioBlob.data);
    }
    showDILink(url);

    //Log dans la console
    console.log("Data_available_!!!");
    console.log(audioBlob);

    //Envoie Ã la console une adresse de tÃ©lÃ©chargement de l'
    Ã©chantillon
    console.log(url);

    //Communique les data au serveur
    servInteract(audioBlob.data, blobType);
}

```

```

function showDlLink(url){
    console.log(url);
}

////////////////////////////////////
////////////////////////////////////
////////////////////////////////////

function mozinitRecording(localMediaStream){
    /* Initialise l'enregistrement */
    mediaRecorder = new MediaRecorder(localMediaStream);
    mediaRecorder.ondataavailable = mozmediaOnDataAvailable;
    mediaStream = localMediaStream;

    console.log('getUserMedia_initialised');
}

function mozstartRecorder(){
    mediaRecorder.start();
    //var audioElement = document.getElementById('audio');
    //audioElement.src = window.URL.createObjectURL(mediaStream);
    //console.log(audioElement.src);
}

function mozstopRecorder(){
    if (mediaStream){
        console.log('stopRecord');
        mediaRecorder.stop();
        //var audioElement = document.getElementById('audio');
        //audioElement.src = '';
    }
}

function mozmediaOnDataAvailable(blob){
    /* A la fin de l'enregistrement, r  cup  re le blob dans data
       et lance le traitement */
    console.log("moz_data_available");
    preInteract(blob, 'ogg');
}

////////////////////////////////////
////////////////////////////////////

function webkitinitRecording(localMediaStream){
    var input = webkitaudio_context.createMediaStreamSource(
        localMediaStream);
    //input.connect(webkitaudio_context.destination);
    webkitrecorder = new Recorder(input);
}

```

```

}

function webkitstartRecorder(){
    webkitrecorder && webkitrecorder.record();
}

function webkitstopRecorder(){
    webkitrecorder && webkitrecorder.stop();

    webkitrecorder && webkitrecorder.exportWAV(function(audioBlob) {
        preInteract(audioBlob, 'wav');
    });

    webkitrecorder.clear();
}

////////////////////
////////////////////
////////////////////

function servInteract(audioBlob, blobType){
    // Envoie le blob au serveur
    var formData = new FormData();
    formData.append('user', user);
    formData.append('hashedPass', hashedPass);
    formData.append('clientDB', clientDB);

    formData.append('action', 'recognize_spoken_word');

    formData.append('audioBlob', audioBlob);
    formData.append('audioType', blobType);

    var req = new XMLHttpRequest();
    req.open('POST', 'handler', false);
    /*req.onreadystatechange = function(){
        console.log('ez');
        console.log(req.readyState);
        if (req.readyState === 4){
            console.log('4');
            if (req.status === 200){
                console.log(200);
                wordResponse(req.responseXML);
            }
        }
    }*/
    //req.setRequestHeader("Content-type", "application/x-www-form-
    urlencoded");
    req.send(formData);
    console.log(req);
    resp = req.responseXML;
    console.log(resp);
}

```

```

        wordResponse(resp);
    }

    function wordResponse(respXML) {
        if (respXML.getElementsByTagName('respWord')) {
            var responseWord = respXML.getElementsByTagName('respWord')[0].
                textContent;
            console.log(responseWord.name);
        }
        else {
            var responseWord = "Error: ";
        }
        var responseElement = document.getElementById('responseWord');
        console.log(responseWord);
        responseElement.innerHTML = responseWord;
    }
}

```

F.2 holder

```

var Holder = Holder || {};
(function (app, win) {

    var preempted = false,
        fallback = false,
        canvas = document.createElement('canvas');
    var dpr = 1, bsr = 1;
    var resizable_images = [];

    if (!canvas.getContext) {
        fallback = true;
    } else {
        if (canvas.toDataURL("image/png")
            .indexOf("data:image/png") < 0) {
            //Android doesn't support data URI
            fallback = true;
        } else {
            var ctx = canvas.getContext("2d");
        }
    }

    if(!fallback){
        dpr = window.devicePixelRatio || 1,
        bsr = ctx.webkitBackingStorePixelRatio || ctx.
            mozBackingStorePixelRatio || ctx.msBackingStorePixelRatio || ctx.
            oBackingStorePixelRatio || ctx.backingStorePixelRatio || 1;
    }

    var ratio = dpr / bsr;

```

```

var settings = {
    domain: "holder.js",
    images: "img",
    bgnodes: ".holderjs",
    themes: {
        "gray": {
            background: "#eee",
            foreground: "#aaa",
            size: 12
        },
        "social": {
            background: "#3a5a97",
            foreground: "#fff",
            size: 12
        },
        "industrial": {
            background: "#434A52",
            foreground: "#C2F200",
            size: 12
        },
        "sky": {
            background: "#0D8FDB",
            foreground: "#fff",
            size: 12
        },
        "vine": {
            background: "#39DBAC",
            foreground: "#1E292C",
            size: 12
        },
        "lava": {
            background: "#F8591A",
            foreground: "#1C2846",
            size: 12
        }
    },
    stylesheet: ""
};

app.flags = {
    dimensions: {
        regex: /^(\d+)x(\d+)$/ ,
        output: function (val) {
            var exec = this.regex.exec(val);
            return {
                width: +exec[1],
                height: +exec[2]
            }
        }
    },
    fluid: {
        regex: /^([0-9%]+)x([0-9%]+)$/ ,
        output: function (val) {
            var exec = this.regex.exec(val);

```

```

        return {
            width: exec[1],
            height: exec[2]
        }
    },
    colors: {
        regex: /#([0-9a-f]{3,})\:#([0-9a-f]{3,})/i,
        output: function (val) {
            var exec = this.regex.exec(val);
            return {
                size: settings.themes.gray.size,
                foreground: "#" + exec[2],
                background: "#" + exec[1]
            }
        }
    },
    text: {
        regex: /text\:(.*)/,
        output: function (val) {
            return this.regex.exec(val)[1];
        }
    },
    font: {
        regex: /font\:(.*)/,
        output: function (val) {
            return this.regex.exec(val)[1];
        }
    },
    auto: {
        regex: /^auto$/
    },
    textmode: {
        regex: /textmode\:(.*)/,
        output: function (val) {
            return this.regex.exec(val)[1];
        }
    }
}

//getElementsByClassName polyfill
document.getElementsByClassName||(document.getElementsByClassName=function(e){var t=document,n,r,i,s=[];if(t.querySelectorAll)return t.querySelectorAll("."+e);if(t.evaluate){r="//*[contains(concat(' ',@class,' '), '"+e+"')]";n=t.evaluate(r,t,null,0,null);while(n.iterateNext())s.push(i)}else{n=t.getElementsByTagName("*"),r=new RegExp("(^|\\s)"+e+"(\\s|$)");for(i=0;i<n.length;i++)r.test(n[i].className)&&s.push(n[i])}return s}}

//getComputedStyle polyfill
window.getComputedStyle||(window.getComputedStyle=function(e){return this.el=e,this.getPropertyValue=function(t){var n=/\-([a-z]){1}/g;return t=="float"&&(t="styleFloat"),n.test(t)&&(t=t.replace(n,function()

```

```

    return arguments[2].toUpperCase())) , e.currentStyle[t]?e.currentStyle[t]:null},this})

//http://javascript.nwbox.com/ContentLoaded by Diego Perini with
    modifications
function contentLoaded(n,t){var l="complete",s="readystatechange",u=!1,h=u,c=!0,i=n.document,a=i.documentElement,e=i.addEventListener?"addEventListener":"attachEvent",v=i.addEventListener?"removeEventListener":"detachEvent",f=i.addEventListener?"":"on",r=function(e){(e.type!=s||i.readyState==l)&&((e.type=="load"?n:i)[v](f+e.type,r,u),!h&&(h=!0)&&t.call(n,null))},o=function(){try{a.doScroll("left")}catch(n){setTimeout(o,50);return}r("poll")};if(i.readyState==l)t.call(n,"lazy");else{if(i.createEventObject&&a.doScroll){try{c=!n.frameElement}catch(y){}}c&&o()}i[e](f+"DOMContentLoaded",r,u),i[e](f+s,r,u),n[e](f+"load",r,u)}}

//https://gist.github.com/991057 by Jed Schmidt with modifications
function selector(a){
    a=a.match(/^(\W)?(.*)/);var b=document["getElement"+(a[1]?a[1]=="#"?"ById":"sByClassName":"sByTagName"])(a[2]);
    var ret=[];    b!=null&&(b.length?ret=b:b.length===0?ret=b:ret=[b]);    return ret;
}

//shallow object property extend
function extend(a,b){
    var c={};
    for(var i in a){
        if(a.hasOwnProperty(i)){
            c[i]=a[i];
        }
    }
    for(var i in b){
        if(b.hasOwnProperty(i)){
            c[i]=b[i];
        }
    }
    return c
}

//hasOwnProperty polyfill
if (!Object.prototype.hasOwnProperty)
    /*jshint -W001, -W103 */
    Object.prototype.hasOwnProperty = function(prop) {
        var proto = this.__proto__ || this.constructor.prototype;
        return (prop in this) && (!(prop in proto) || proto[prop] !== this[prop]);
    }
    /*jshint +W001, +W103 */

function text_size(width, height, template) {
    height = parseInt(height, 10);
    width = parseInt(width, 10);

```

```

    var bigSide = Math.max(height, width)
    var smallSide = Math.min(height, width)
    var scale = 1 / 12;
    var newHeight = Math.min(smallSide * 0.75, 0.75 * bigSide * scale)
    ;
    return {
        height: Math.round(Math.max(template.size, newHeight))
    }
}

function draw(args) {
    var ctx = args.ctx;
    var dimensions = args.dimensions;
    var template = args.template;
    var ratio = args.ratio;
    var holder = args.holder;
    var literal = holder.textmode == "literal";
    var exact = holder.textmode == "exact";

    var ts = text_size(dimensions.width, dimensions.height, template);
    var text_height = ts.height;
    var width = dimensions.width * ratio,
        height = dimensions.height * ratio;
    var font = template.font ? template.font : "sans-serif";
    canvas.width = width;
    canvas.height = height;
    ctx.textAlign = "center";
    ctx.textBaseline = "middle";
    ctx.fillStyle = template.background;
    ctx.fillRect(0, 0, width, height);
    ctx.fillStyle = template.foreground;
    ctx.font = "bold_" + text_height + "px_" + font;
    var text = template.text ? template.text : (Math.floor(dimensions.
        width) + "x" + Math.floor(dimensions.height));
    if (literal) {
        var dimensions = holder.dimensions;
        text = dimensions.width + "x" + dimensions.height;
    }
    else if(exact && holder.exact_dimensions){
        var dimensions = holder.exact_dimensions;
        text = (Math.floor(dimensions.width) + "x" + Math.floor(
            dimensions.height));
    }
    var text_width = ctx.measureText(text).width;
    if (text_width / width >= 0.75) {
        text_height = Math.floor(text_height * 0.75 * (width /
            text_width));
    }
    //Resetting font size if necessary
    ctx.font = "bold_" + (text_height * ratio) + "px_" + font;
    ctx.fillText(text, (width / 2), (height / 2), width);
    return canvas.toDataURL("image/png");
}

```



```

function render(mode, el, holder, src) {

    var dimensions = holder.dimensions,
        theme = holder.theme,
        text = holder.text ? decodeURIComponent(holder.text) :
            holder.text;
    var dimensions_caption = dimensions.width + "x" + dimensions.
        height;
    theme = (text ? extend(theme, {
        text: text
    }) : theme);
    theme = (holder.font ? extend(theme, {
        font: holder.font
    }) : theme);
    el.setAttribute("data-src", src);
    holder.theme = theme;
    el.holder_data = holder;

    if (mode == "image") {
        el.setAttribute("alt", text ? text : theme.text ? theme.
            text + "┐[" + dimensions_caption + "]" :
            dimensions_caption);
        if (fallback || !holder.auto) {
            el.style.width = dimensions.width + "px";
            el.style.height = dimensions.height + "px";
        }
        if (fallback) {
            el.style.backgroundColor = theme.background;
        } else {
            el.setAttribute("src", draw({ctx: ctx, dimensions:
                dimensions, template: theme, ratio: ratio,
                holder: holder}));

            if(holder.textmode && holder.textmode == "exact"){
                resizable_images.push(el);
                resizable_update(el);
            }
        }
    }
    else if (mode == "background") {
        if (!fallback) {
            el.style.backgroundImage = "url(" + draw({ctx: ctx,
                dimensions: dimensions, template: theme,
                ratio: ratio, holder: holder}) + ")";
            el.style.backgroundSize = dimensions.width + "px┐"
                + dimensions.height + "px";
        }
    }
    else if (mode == "fluid") {
        el.setAttribute("alt", text ? text : theme.text ? theme.
            text + "┐[" + dimensions_caption + "]" :
            dimensions_caption);
        if (dimensions.height.slice(-1) == "%") {

```

```

        el.style.height = dimensions.height
    } else {
        el.style.height = dimensions.height + "px"
    }
    if (dimensions.width.slice(-1) === "%") {
        el.style.width = dimensions.width
    } else {
        el.style.width = dimensions.width + "px"
    }
    if (el.style.display === "inline" || el.style.display === "
" || el.style.display === "none") {
        el.style.display = "block";
    }
    if (fallback) {
        el.style.backgroundColor = theme.background;
    } else {
        resizable_images.push(el);
        resizable_update(el);
    }
}
}

function dimension_check(el, callback) {
    var dimensions = {
        height: el.clientHeight,
        width: el.clientWidth
    };
    if (!dimensions.height && !dimensions.width) {
        if (el.hasAttribute("data-holder-invisible")) {
            throw new Error("Holder:_placeholder_is_not_
            visible");
        } else {
            el.setAttribute("data-holder-invisible", true)
            setTimeout(function () {
                callback.call(this, el)
            }, 1)
            return null;
        }
    } else {
        el.removeAttribute("data-holder-invisible")
    }
    return dimensions;
}

function resizable_update(element) {
    var images;
    if (element.nodeType === null) {
        images = resizable_images;
    } else {
        images = [element]
    }
    for (var i in images) {
        if (!images.hasOwnProperty(i)) {

```

```

        continue;
    }
    var el = images[i]
    if (el.holder_data) {
        var holder = el.holder_data;
        var dimensions = dimension_check(el,
            resizable_update)
        if(dimensions){
            if(holder.fluid){
                el.setAttribute("src", draw({
                    ctx: ctx,
                    dimensions: dimensions,
                    template: holder.theme,
                    ratio: ratio,
                    holder: holder
                }))
            }
            if(holder.textmode && holder.textmode == "
                exact"){
                holder.exact_dimensions =
                    dimensions;
                el.setAttribute("src", draw({
                    ctx: ctx,
                    dimensions: holder.
                        dimensions,
                    template: holder.theme,
                    ratio: ratio,
                    holder: holder
                }))
            }
        }
    }
}

function parse_flags(flags, options) {
    var ret = {
        theme: extend(settings.themes.gray, {})
    };
    var render = false;
    for (sl = flags.length, j = 0; j < sl; j++) {
        var flag = flags[j];
        if (app.flags.dimensions.match(flag)) {
            render = true;
            ret.dimensions = app.flags.dimensions.output(flag)
        }
        else if (app.flags.fluid.match(flag)) {
            render = true;
            ret.dimensions = app.flags.fluid.output(flag);
            ret.fluid = true;
        }
        else if (app.flags.textmode.match(flag)) {
            ret.textmode = app.flags.textmode.output(flag)
        }
        else if (app.flags.colors.match(flag)) {

```

```

        ret.theme = app.flags.colors.output(flag);
    } else if (options.themes[flag]) {
        //If a theme is specified, it will override custom
        colors
        if(options.themes.hasOwnProperty(flag)){
            ret.theme = extend(options.themes[flag],
                                {});
        }
    } else if (app.flags.font.match(flag)) {
        ret.font = app.flags.font.output(flag);
    } else if (app.flags.auto.match(flag)) {
        ret.auto = true;
    } else if (app.flags.text.match(flag)) {
        ret.text = app.flags.text.output(flag);
    }
}
return render ? ret : false;
}

for (var flag in app.flags) {
    if (!app.flags.hasOwnProperty(flag)) continue;
    app.flags[flag].match = function (val) {
        return val.match(this.regex)
    }
}

app.add_theme = function (name, theme) {
    name != null && theme != null && (settings.themes[name] = theme);
    return app;
};

app.add_image = function (src, el) {
    var node = selector(el);
    if (node.length) {
        for (var i = 0, l = node.length; i < l; i++) {
            var img = document.createElement("img")
            img.setAttribute("data-src", src);
            node[i].appendChild(img);
        }
    }
    return app;
};

app.run = function (o) {
    preempted = true;

    var options = extend(settings, o),
        images = [],
        imageNodes = [],
        bgnodes = [];
    if (typeof (options.images) == "string") {
        imageNodes = selector(options.images);
    } else if (window.NodeList && options.images instanceof window.
        NodeList) {
        imageNodes = options.images;
    } else if (window.Node && options.images instanceof window.Node) {

```

```

        imageNodes = [options.images];
    }

    if (typeof (options.bgnodes) == "string") {
        bgnodes = selector(options.bgnodes);
    } else if (window.NodeList && options.elements instanceof window.
        NodeList) {
        bgnodes = options.bgnodes;
    } else if (window.Node && options.bgnodes instanceof window.Node)
    {
        bgnodes = [options.bgnodes];
    }
    for (i = 0, l = imageNodes.length; i < l; i++) images.push(
        imageNodes[i]);
    var holdercss = document.getElementById("holderjs-style");
    if (!holdercss) {
        holdercss = document.createElement("style");
        holdercss.setAttribute("id", "holderjs-style");
        holdercss.type = "text/css";
        document.getElementsByTagName("head")[0].appendChild(
            holdercss);
    }
    if (!options.nocss) {
        if (holdercss.styleSheet) {
            holdercss.styleSheet.cssText += options.stylesheet
                ;
        } else {
            holdercss.appendChild(document.createTextNode(
                options.stylesheet));
        }
    }
    var cssregex = new RegExp(options.domain + "\\/(.*?)\\"?\\");
    for (var l = bgnodes.length, i = 0; i < l; i++) {
        var src = window.getComputedStyle(bgnodes[i], null)
            .getPropertyValue("background-image");
        var flags = src.match(cssregex);
        var bgsrc = bgnodes[i].getAttribute("data-background-src")
            ;
        if (flags) {
            var holder = parse_flags(flags[1].split("/"),
                options);
            if (holder) {
                render("background", bgnodes[i], holder,
                    src);
            }
        } else if (bgsrc != null) {
            var holder = parse_flags(bgsrc.substr(bgsrc.
                lastIndexOf(options.domain) + options.domain.
                length + 1)
                .split("/"), options);
            if (holder) {
                render("background", bgnodes[i], holder,
                    src);
            }
        }
    }

```

```

        }
    }
}
for (l = images.length, i = 0; i < l; i++) {
    var attr_data_src, attr_src;
    attr_src = attr_data_src = src = null;
    try {
        attr_src = images[i].getAttribute("src");
        attr_datasrc = images[i].getAttribute("data-src");
    } catch (e) {}
    if (attr_datasrc == null && !! attr_src && attr_src.
        indexOf(options.domain) >= 0) {
        src = attr_src;
    } else if ( !! attr_datasrc && attr_datasrc.indexOf(
        options.domain) >= 0) {
        src = attr_datasrc;
    }
    if (src) {
        var holder = parse_flags(src.substr(src.
            lastIndexOf(options.domain) + options.domain.
            length + 1)
            .split("/"), options);
        if (holder) {
            if (holder.fluid) {
                render("fluid", images[i], holder,
                    src)
            } else {
                render("image", images[i], holder,
                    src);
            }
        }
    }
}
}
return app;
};
contentLoaded(win, function () {
    if (window.addEventListener) {
        window.addEventListener("resize", resizable_update, false)
        ;
        window.addEventListener("orientationchange",
            resizable_update, false);
    } else {
        window.attachEvent("onresize", resizable_update)
    }
    preempted || app.run();
});
if (typeof define == "function" && define.amd) {
    define([], function () {
        return app;
    });
}
})(Holder, window);

```

F.3 recorder

```
(function(window){

  var WORKER_PATH = 'js/recorderWorker.js';

  var Recorder = function(source, cfg){
    var config = cfg || {};
    var bufferLen = config.bufferLen || 4096;
    this.context = source.context;
    this.node = this.context.createJavaScriptNode(bufferLen, 2, 2);
    var worker = new Worker(config.workerPath || WORKER_PATH);
    worker.postMessage({
      command: 'init',
      config: {
        sampleRate: this.context.sampleRate
      }
    });
    var recording = false,
        currCallback;

    this.node.onaudioprocess = function(e){
      if (!recording) return;
      worker.postMessage({
        command: 'record',
        buffer: [
          e.inputBuffer.getChannelData(0),
          e.inputBuffer.getChannelData(1)
        ]
      });
    };

    this.configure = function(cfg){
      for (var prop in cfg){
        if (cfg.hasOwnProperty(prop)){
          config[prop] = cfg[prop];
        }
      }
    };

    this.record = function(){
      recording = true;
    };

    this.stop = function(){
      recording = false;
    };

    this.clear = function(){
      worker.postMessage({ command: 'clear' });
    };

    this.getBuffer = function(cb) {
```

```

    currCallback = cb || config.callback;
    worker.postMessage({ command: 'getBuffer' })
}

this.exportWAV = function(cb, type){
    currCallback = cb || config.callback;
    type = type || config.type || 'audio/wav';
    if (!currCallback) throw new Error('Callback_not_set');
    worker.postMessage({
        command: 'exportWAV',
        type: type
    });
}

worker.onmessage = function(e){
    var blob = e.data;
    currCallback(blob);
}

source.connect(this.node);
//this.node.connect(this.context.destination);    //this should not be
//necessary
};

Recorder.forceDownload = function(blob, filename){
    var url = (window.URL || window.webkitURL).createObjectURL(blob);
    var link = window.document.createElement('a');
    link.href = url;
    link.download = filename || 'output.wav';
    var click = document.createEvent("Event");
    click.initEvent("click", true, true);
    link.dispatchEvent(click);
}

window.Recorder = Recorder;
})(window);

```

F.4 recorderWorker

```

var recLength = 0,
    recBuffersL = [],
    recBuffersR = [],
    sampleRate;

this.onmessage = function(e){
    switch(e.data.command){
        case 'init':
            init(e.data.config);
            break;
        case 'record':
            record(e.data.buffer);

```



```

        break;
    case 'exportWAV':
        exportWAV(e.data.type);
        break;
    case 'getBuffer':
        getBuffer();
        break;
    case 'clear':
        clear();
        break;
    }
};

function init(config){
    sampleRate = config.sampleRate;
}

function record(inputBuffer){
    recBuffersL.push(inputBuffer[0]);
    recBuffersR.push(inputBuffer[1]);
    recLength += inputBuffer[0].length;
}

function exportWAV(type){
    var bufferL = mergeBuffers(recBuffersL, recLength);
    var bufferR = mergeBuffers(recBuffersR, recLength);
    var interleaved = interleave(bufferL, bufferR);
    var dataview = encodeWAV(interleaved);
    var audioBlob = new Blob([dataview], { type: type });

    this.postMessage(audioBlob);
}

function getBuffer() {
    var buffers = [];
    buffers.push(mergeBuffers(recBuffersL, recLength));
    buffers.push(mergeBuffers(recBuffersR, recLength));
    this.postMessage(buffers);
}

function clear(){
    recLength = 0;
    recBuffersL = [];
    recBuffersR = [];
}

function mergeBuffers(recBuffers, recLength){
    var result = new Float32Array(recLength);
    var offset = 0;
    for (var i = 0; i < recBuffers.length; i++){
        result.set(recBuffers[i], offset);
        offset += recBuffers[i].length;
    }
}

```

```

    return result;
}

function interleave(inputL, inputR){
    var length = inputL.length + inputR.length;
    var result = new Float32Array(length);

    var index = 0,
        inputIndex = 0;

    while (index < length){
        result[index++] = inputL[inputIndex];
        result[index++] = inputR[inputIndex];
        inputIndex++;
    }
    return result;
}

function floatTo16BitPCM(output, offset, input){
    for (var i = 0; i < input.length; i++, offset+=2){
        var s = Math.max(-1, Math.min(1, input[i]));
        output.setInt16(offset, s < 0 ? s * 0x8000 : s * 0x7FFF, true);
    }
}

function writeString(view, offset, string){
    for (var i = 0; i < string.length; i++){
        view.setUint8(offset + i, string.charCodeAt(i));
    }
}

function encodeWAV(samples){
    var buffer = new ArrayBuffer(44 + samples.length * 2);
    var view = new DataView(buffer);

    /* RIFF identifier */
    writeString(view, 0, 'RIFF');
    /* file length */
    view.setUint32(4, 32 + samples.length * 2, true);
    /* RIFF type */
    writeString(view, 8, 'WAVE');
    /* format chunk identifier */
    writeString(view, 12, 'fmt ');
    /* format chunk length */
    view.setUint32(16, 16, true);
    /* sample format (raw) */
    view.setUint16(20, 1, true);
    /* channel count */
    view.setUint16(22, 2, true);
    /* sample rate */
    view.setUint32(24, sampleRate, true);
    /* byte rate (sample rate * block align) */
    view.setUint32(28, sampleRate * 4, true);

```

```

    /* block align (channel count * bytes per sample) */
    view.setUint16(32, 4, true);
    /* bits per sample */
    view.setUint16(34, 16, true);
    /* data chunk identifier */
    writeString(view, 36, 'data');
    /* data chunk length */
    view.setUint32(40, samples.length * 2, true);

    floatTo16BitPCM(view, 44, samples);

    return view;
}

```

F.5 index.html

```

<!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="utf-8">
    <meta http-equiv="X-UA-Compatible" content="IE=edge">
    <meta name="viewport" content="width=device-width, _initial-scale=1.0">
    <meta name="description" content="MIG_SE_Speech_Recognition_
      Demonstrator">
    <meta name="author" content="MIG_SE_Team">
    <link rel="shortcut_icon" href="img/favicon.png">

    <title>SpeechApp demonstrator</title>

    <!-- Bootstrap core CSS -->
    <link href="css/bootstrap.min.css" rel="stylesheet">
    <!-- Bootstrap theme -->
    <link href="css/bootstrap-theme.min.css" rel="stylesheet">

    <!-- Custom styles for this template -->
    <link href="css/theme.css" rel="stylesheet">

    <!-- Just for debugging purposes. Don't actually copy this line! -->
    <!--[if lt IE 9]><script src="../../docs-assets/js/ie8-responsive-file
      -warning.js"></script><![endif]-->

    <!-- HTML5 shim and Respond.js IE8 support of HTML5 elements and media
      queries -->
    <!--[if lt IE 9]>
      <script src="https://oss.maxcdn.com/libs/html5shiv/3.7.0/html5shiv.
        js"></script>
      <script src="https://oss.maxcdn.com/libs/respond.js/1.3.0/respond.
        min.js"></script>
    <![endif]-->
  </head>

  <body>

```

```

<!-- Fixed navbar -->
<div class="navbar navbar-inverse navbar-fixed-top" role="navigation">
  <div class="container">
    <div class="navbar-header">
      <button type="button" class="navbar-toggle" data-toggle="
        collapse" data-target=".navbar-collapse">
        <span class="sr-only">Toggle navigation</span>
        <span class="icon-bar"></span>
        <span class="icon-bar"></span>
        <span class="icon-bar"></span>
      </button>
      <a class="navbar-brand" href="#">SpeechApp</a>
    </div>
    <div class="navbar-collapse collapse">
      <ul class="nav navbar-nav">
        <li class="active"><a href="#">Home</a></li>
        <li><a href="#services">Our services</a></li>
        <li class="dropdown">
          <a href="#" class="dropdown-toggle" data-toggle="dropdown">
            About <b class="caret"></b></a>
          <ul class="dropdown-menu">
            <li><a href="#team">Our team</a></li>
            <li><a href="#work">Our work</a></li>
            <li><a href="#contact">Contact</a></li>
          </ul>
        </li>
      </ul>
    </div>
  </div>
</div>

<div class="container theme-showcase">

  <!-- Main jumbotron for a primary marketing message or call to
    action -->
  <div class="jumbotron">
    <h1>SpeechApp !</h1>
    <p>Welcome to this demonstrator of our state of the art speech
      recognition technology</p>
    <p>Test it, feed us with a few words</p>
    <p><a href="#services" class="btn btn-primary btn-lg" role="button
      ">Learn more &raquo;</a></p>
  </div>

  <div class="page-header" id="demonstrator">
    <h1>Demonstrator</h1>
  </div>

  <p>
    <div style="text-align:center;" id="sound-recording">

```

```

    <p id="recording">
      <audio autoplay src="" id="audio">audio</audio>
    </p>
    <p id="responseWord" style="color:_blue" ></p>
  </div>
</p>

<div class="page-header" id="services">
  <h1>Our services</h1>
</div>

<p>

</p>

<div class="page-header" id="team">
  <h1>Our team</h1>
</div>

<p>
  <p style="text-align:_center;">
    
  </p>
  We are a team of 13 first year students at <a href="http://www.
    mines-paristech.eu/">MINES ParisTech</a><br />
</p>

<div class="page-header" id="work">
  <h1>Our work</h1>
</div>

<p>
  We've been working for 3 weeks on speech recognition at the <a
    href="http://www.cma.ensmp.fr/">CMA</a>. We've build from
    scratch an isolated
    spoken word recognition engine that works pretty well, and the
    tools to use it easily. This webapp is
    part of it.
</p>
<p>Our project has been written with Python, and its source is
  hosted on
  <a href="https://github.com/giliam/mig2013">our Git repository</a>
</p>

<div class="page-header" id="contact">
  <h1>Contact</h1>
</div>

<p>

```

```

    You'll find our GitHub profiles on
    <a href="https://github.com/giliam/mig2013">our Git repository
    page</a>.<br />
    The maintainer of that webapp can be contacted at
    <a href="mailto:speechapp@wumzi.info">speechapp@wumzi.info</a>.
</p>

```

```
</div> <!-- /container -->
```

```

<!-- SpeechApp js core -->
<script src="js/main.js"></script>
<script src="js/recorder.js"></script>

```

```
<!-- Bootstrap core JavaScript
```

```
===== -->
```

```
<!-- Placed at the end of the document so the pages load faster -->
```

```
<script src="js/jquery-1.10.2.min.js"></script>
```

```
<script src="js/bootstrap.min.js"></script>
```

```
<script src="js/holder.js"></script>
```

```
</body>
```

```
</html>
```

G. SpeechServer

G.1 main.py

```

class SpeechServerHandler( BaseHTTPServer.BaseHTTPRequestHandler ):
    def do_GET(self):
        """ Respond to a GET request """
        self.send_response(200)
        self.send_header( 'Content-type', 'text/plain' )
        self.end_headers()
        self.wfile.write("Ca_se_passe_en_POST_pour_les_requetes!")

    def do_POST(self):
        """Respond to a POST request"""

        form = FieldStorage(
            fp=self.rfile,
            headers=self.headers,
            environ={ 'REQUEST_METHOD': 'POST',
                      'CONTENT_TYPE': self.headers[ 'Content-Type' ],
                    })

```

```

user = form.getvalue('user')
hashedPass = form.getvalue('hashedPass')
clientDB = ''#form.getvalue('clientDB')

#form = dict(form)
#print(form)

#Check if the user is authorized and he has access to clientDb
authUser = AuthUser()
if True or authUser.checkAuth(user, authUser.hashPass(hashedPass),
    clientDB):
    action = form.getvalue('action')
    requestHandler = requestHandling()
    respData = requestHandler.handle(clientDB, action, form)
    respXML = self.buildXMLResponse(respData)
else:
    respXML = "You're_not_authorized_to_call_me!\n
    ~~~~~~Register_at_speech.wumzi.info"

#respXML = self.buildXMLResponse({'respWord' : user})

#And respond
self.send_response(200)
self.send_header('Content-type', 'text/xml')
self.end_headers()
self.wfile.write(respXML)

@classmethod
def buildXMLResponse(cls, data):
    """Build the XML doc response from the data dictionary"""
    root = ET.Element('root')
    for key, value in data.items():
        elem = ET.SubElement(root, key)
        elem.text = value

    return ET.tostring(root, encoding="utf-8")

@classmethod
def parseXMLRequest(cls, XMLString):
    """Build a dict from an XML doc"""
    data = {}
    root = ET.fromstring(XMLString)
    for child in root:
        data[child.tag] = child.text
    return data

def run(port, adress="localhost"):
    server = BaseHTTPServer.HTTPServer((adress, port),
        SpeechServerHandler)

```

```

server.serve_forever()

if __name__ == '__main__':
    import sys
    if len(sys.argv) >= 2:
        try:
            PORT = int(sys.argv[1])
        except TypeError:
            print("Please_provide_an_int_!")
    else:
        PORT = 8010
        print("Port_set_to_default_:_%s" % PORT)

run('localhost', PORT)

```

G.2 audioConverter.py

```

TMP_DIR = "tmp/"

def path_orig_ogg(id):
    return TMP_DIR + "orig_" + str(id) + ".ogg"

def path_mid_wave(id):
    return TMP_DIR + "orig_" + str(id) + ".wav"

def path_mid_waveSplitted(id):
    return TMP_DIR, "orig_" + str(id) + ".wav"

def path_final_wave(id):
    return TMP_DIR + "wave_final_" + str(id) + ".wav"

def path_final_waveSplitted(id):
    return TMP_DIR, "wave_final_" + str(id) + ".wav"

def rm_multi(*files):
    """Remove multiple files"""
    for path in files:
        os.remove(path)

def handleOGGBlob(oggBlob):
    """Converti le blob ogg en blob wav"""
    id = randint(1, 1000)
    while os.access(path_orig_ogg(id), os.W_OK):
        id = randint(1, 1000)

    writeBlobToDisk(oggBlob, path_orig_ogg(id))

    #Now convert the file

    print(path_orig_ogg(id))

```



```

os.system('soundconverter -b -m audio/x-wav -s %.wav "%s"' %
    path_orig_ogg(id))
#Resample to 44.1kHz

return finalHandling(id)

def finalHandling(id):
    print('final_handling_started')
    os.system('sox -r 44.1k -e signed -c 1 -b 16 %s %s' % (path_mid_wave(
        id), path_final_wave(id)))
    print('soxed')
    #And read the oggBlob

    '''with open(path_final_wave(id), 'r') as finalwavefile:
        waveBlob = finalwavefile.read()'''

    dir, file = path_final_wave_split(id)
    print(dir, file)
    waveBlob = cutsyncaudio(dir, file)

    #Remove the files
    rm_multi(path_mid_wave(id), )#path_final_wave(id)
    print("success")
    return waveBlob

def handleWAVBlob(audioBlob):
    id = randint(1, 1000)
    while os.access(path_mid_wave(id), os.W_OK):
        id = randint(1, 1000)

    writeBlobToDisk(audioBlob, path_mid_wave(id))
    print path_mid_wave(id)
    print "bringing_id_to_finalHandling"
    return finalHandling(id)

def writeBlobToDisk(audioBlob, path):
    with open(path, 'w') as origfile:
        origfile.write(audioBlob)
        origfile.flush()
        os.fsync(origfile)

def cutsyncaudio(dir, file):
    sync.cutBeginning(dir, file, prefix='')
    sync.syncFile(dir, file, prefix='')

    waveBlob = scipy.io.wavfile.read(dir + file)

    return waveBlob

```

```

def convert_ogg_to_wav(ogg_path, out_wav_path):
    try:
        with open(ogg_path, 'r') as origoggfile:
            waveBlob = handleOGGBlob(origoggfile.read())
    except:
        return "Impossible_to_open_ogg_file"

    try:
        with open(out_wav_path, 'w') as out_wav:
            out_wav.write(waveBlob)
            out_wav.flush()
            os.fsync(out_wav)
    except:
        return "Impossible_to_write_wav_blob_to_dest_file"

    return waveBlob

def sox_handling(wavBlob, pathToTmp="../db/waves/tmp/"):
    tempFileName = hashlib.sha224(str(randint(0,1e10))).hexdigest()
    fileName = pathToTmp + str(tempFileName) + ".wav"
    with open(fileName, 'w') as origoggfile:
        origoggfile.write(wavBlob)
        origoggfile.flush()
        os.fsync(origoggfile)

    os.system('ffmpeg -i "' + fileName + '" -vn -ss 00:00:00 -t 00:00:01 -'
        + pathToTmp + 'noiseaud.wav')
    print("noise_extracted")
    os.system('sox "' + pathToTmp + 'noiseaud.wav" -n noiseprof "' +
        pathToTmp + 'noise.prof"')
    print("noise_selected")
    sleep(1)
    os.system('sox "' + fileName + '" -n "' + fileName + '" -noisered "' +
        pathToTmp + 'noise.prof" 0.21')
    print("noise_trashed")
    #os.remove(pathToTmp + "noise.prof")
    #os.remove(pathToTmp + "noiseaud.wav")

    wav_content = scipy.io.wavfile.read(fileName)
    return wav_content

if __name__ == '__main__':
    print(sox_handling(convert_ogg_to_wav('test.oga', 'test.wav')))
```

G.3 clientAuth.py

```

DEBUG = False

class AuthUser:
```

```

""" Classe pour g  rer l'authentification des applications clientes
sur notre
serveur applicatif """

def __init__(self, fileName="registre"):
    self.userListFile = fileName
    self.db = Db("../db/", "userDbList", DEBUG)
    self.userList = self.db.getFile("users/" + fileName + ".txt")
    self.username = ""
    self.password = ""
    self.connected = False

def newClient(self, client, hashedPass, authorizedDBs):
    """ Ajoute un utilisateur et des donn  es """
    if not self.getClient(client):
        self.userList[client] = hashedPass, authorizedDBs
        self.commit()
        return True
    return False

def updateClient(self, client, hashedPass, authorizedDBs):
    """ Ajoute un utilisateur et des donn  es """
    if self.userList.get(client):
        self.userList[client] = [hashedPass, authorizedDBs]
        self.commit()
        return True
    return False

def rmClient(self, client):
    """ Supprime un client du dictionnaire """
    if self.userList.get(client):
        del self.userList[client]
        self.commit()
        return True
    return False

def getClients(self):
    print self.userList

def getClient(self, client):
    """ Retourne un client s'il se trouve dans la liste des
    utilisateurs """
    if self.userList.get(client):
        return self.userList[client]
    else:
        return False

def checkAuth(self, client, submittedHashedPass, clientDB=""):

```

```

        """ VÃ©rifie que le nom entrÃ© se trouve bien dans la liste des
        utilisateurs """
    if self.getClient(client):
        hashedPass, clientDBs = self.getClient(client)
        if hashedPass == submittedHashedPass:
            if clientDB == "" or clientDB in clientDBs:
                return True
    return False

def logIn(self, client, submittedHashedPass):
    """ Connecte l'utilisateur """
    if self.getClient(client):
        hashedPass, clientDBs = self.getClient(client)
        if hashedPass == submittedHashedPass:
            self.username = client
            self.password = submittedHashedPass
            self.connected = True
            return True
    return False

def logOut(self):
    self.username = ""
    self.password = ""
    self.connected = False

def hashPass(self, password):
    return hashlib.sha224(password).hexdigest()

def commit(self):
    """ Write the changes of the userlist to the DB on disk """
    self.db.addFile("users/" + self.userListFile + ".txt", self.
        userList)

def __str__(self):
    """ Affiche la liste des utilisateurs et leurs donnÃ©es """
    data = ["%s_:\t%s" % (client, clientData) for client, clientData
        in self.userList.items()]
    return '\n'.join(data)

if __name__ == "__main__":
    authUserHandler = AuthUser()
    print(authUserHandler)
    authUserHandler.newClient("demo", "demo", [1])
    print(authUserHandler)

```

G.4 speechActions.py

```

actions possibles : add_word_record, list_word_records, rm_word_record,
                    recognize_spoken_word """

```

```

from shell import *
from core.utils.db import Db
from core.recording import sync #import syncFile, cutBeginning

from speechserver.audioConverter import *

ACTIONS = ["add_word", "list_word_records", "rm_word_record", "
    recognize_spoken_word", "listen_recording"]

class requestHandling:
    def handle(self, clientDBid, action, data):
        self.dbWaves = Db('../db/', verbose=False)

        if not action in ACTIONS:
            return False

        if action == "recognize_spoken_word":
            audioBlob = data.getvalue("audioBlob")
            audioType = data.getvalue("audioType")
            if not (audioBlob or audioType):
                return None
            else:
                return self.recognize_spoken_word(audioBlob, audioType,
                    clientDBid)

    def recognize_spoken_word(self, audioBlob, audioType, clientDBid):
        """ Handle the specific request to recognize a spoken word """
        TYPES = ['wav', 'ogg']
        if audioType not in TYPES:
            return False

        if audioType == 'ogg':
            audioBlob = handleOGGBlob(audioBlob)
            print(audioBlob)
        elif audioType == 'wav':
            audioBlob = handleWAVBlob(audioBlob)

        #wav_content = sox_handling(audioBlob)
        #print wav_content
        respWord, log = handlingOneWord(audioBlob[1], self.dbWaves, 1, 1,
            0)
        print respWord
        return {'respWord': respWord}

```

Bibliographie

- [1] Apple. Application siri. <http://www.apple.com/fr/ios/siri/>, 2013.
- [2] Tom Preston-Werner, Chris Wanstrath, and PJ Hyett. Github. <http://www.github.com/>, 2013.
- [3] Lawrence Rabiner. *Fundamentals of Speech Recognition*. Prentice Hall PTR, 1993.
- [4] MIT. Pyaudio. <http://people.csail.mit.edu/hubert/pyaudio/>, 2006.
- [5] Stanley Smith Stevens, John Volkman, and Edwin B. Newman. A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America*, 1937.
- [6] Begam, Elamvazuthi, and Muda. Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw). *Journal of Computing*, 2010.
- [7] Vincent Arsigny. Modélisation par un champ de markov du signal de parole et application à la reconnaissance vocale. Technical report, École Nationale Supérieure des Télécom de Paris, 2000.
- [8] Zohar Babin. How to do noise reduction using ffmpeg and sox. <http://www.zoharbabin.com/how-to-do-noise-reduction-using-ffmpeg-and-sox/>, 2011.
- [9] Chris Bagwell. Sox website. <http://sox.sourceforge.net/Docs/Documentation>, 2009.
- [10] H.G. Hirsch, P Meyer, and H.W. Ruehl. Improved speech recognition using high-pass filtering of subband envelopes. *Eurospeech*, 1991.
- [11] Anonyme. Théorème de nyquist-shannon. http://fr.wikipedia.org/wiki/Th%C3%A9or%C3%A8me_d%27%C3%A9chantillonnage_de_Nyquist-Shannon, 2013.
- [12] Luc Maranget. *Introduction à la programmation*. École Polytechnique, 2008-2009.
- [13] Maurice Charbit. Reconnaissance de mots isolés (utilisation des modèles HMM). *Inconnu*, Oct. 2002.
- [14] Franck Bonnet, Benjamin Devèze, Mathieu Fouquin, and Julien Jeany. Reconnaissance automatique de la parole. *Epita*, 2004.
- [15] Aharon Etengoff. Nuance clinches speech-recognition deal with us army. <http://www.itexaminer.com/nuance-clinches-speech-recognition-deal-with-us-army.aspx>, 2009.
- [16] Thanassis Trikas. Automated speech recognition in air traffic control. *MIT*, 1987.
- [17] G2 Speech. La reconnaissance vocale pour hôpitaux et autres institutions de soins. <http://www.g2speech.be/reconnaissance-vocale.html>, 2000.
- [18] Nuance Company. Dragon naturallyspeaking for law enforcement. <http://www.nuance.com/naturallyspeaking/industries/law-enforcement/>, ?
- [19] Nancy Manasse. Speech recognition. *University of Nebraska-Lincoln*, 1990.
- [20] PrismaMedia. Capital. www.capital.fr, ?
- [21] Michael Page. Hays, officeteam, 2009-2010.
- [22] Gouvernement français. site des impôts. www.impots.gouv.fr, 2013.
- [23] Nuance Company. Dragon naturallyspeaking. <http://www.nuance.com/dragon/index.htm>, ?