

[MIG] Systèmes Embarqués

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

Table des matières

0.1	Présentation des enjeux	2
0.2	Objectifs du projet	2
I	Démarche Technique	4
1	Analyse, formatage du signal	4
1.1	Introduction	4
1.2	Prérequis	4
1.2.1	Qu'est-ce que le son ?	4
1.2.2	Comment le son est-il représenté dans l'ordinateur ?	4
1.2.3	Pré accentuation des aigus	5
1.2.4	Etalonnage des fréquences	5
1.3	Enregistrement, recadrage, filtrage HF	5
1.3.1	Synchronisation	5
1.4	Echantillonnage, fenêtrage	6
1.4.1	début du filtre 2	6
1.4.2	fin du filtre 2	6
1.5	Transformée de Fourier	6
1.6	MEL	6
1.7	Transformée inverse	6
2	Les MMC	7
2.1	Principe, modèle discret	7
2.2	Phase d'apprentissage	7
2.3	Phase de reconnaissance	7
II	Approche commerciale	8
1	Côté web	8
2	Applications	9
3	Budget, modèle économique	11
III	Le Code	12
1	Python	12
2	C/C++	14

Introduction

0.1 Présentation des enjeux

La reconnaissance vocale automatisée est l'objet d'intenses recherches depuis plus de 50 ans. Nous apparaissant encore aujourd'hui comme futuriste malgré son omniprésence dans les systèmes embarqués de dernier cri (smartphones, automobile ...), l'aspiration à un système automatisé avec lequel nous pourrions communiquer et même dialoguer est profondément ancré dans l'imaginaire commun, et entretenu par de nombreux films de science fiction, tels l'odyssée Star Wars de George Lucas et tend à devenir une réalité (iPhone). Les perspectives économiques s'ouvrant 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 au sourds/muets, ...) et s'inscrit dans le domaine prolifique beaucoup plus large du traitement du signal. 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, ...) et de part l'approche unidirectionnelle de certaines recherches. 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, si il est stressé ou calme, dans quel environnement est-il, 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 limité imparti. Il ne s'agissait dès lors 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. La complexité de la discipline fut un des principaux obstacles, et une phase d'appropriation des techniques requises, de part la lecture de livre 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. Le projet des MIG ne 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.

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 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. Ceci a permis à chacun d'être en permanence en totale connaissance des rôles de chacun, et de la ligne directrice de chaque sous-fonction sans pour autant en connaître tous les détails. Une efficacité dans la coordination des tâches et dans leur réalisation en a indiscutablement découlé.

Première partie

Démarche Technique

1. Analyse, formatage du signal

1.1 Introduction

Comme nous l'avons mentionné, même le plus élémentaires 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 s'absoudre 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.

1.2 Prérequis

1.2.1 Qu'est-ce que le son ?

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 compose et qui changent au cours du temps.

1.2.2 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. 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 stockée de manière précise, et donc plus la qualité du son sera bonne. Grâce à l'échantillonnage et à, on définit aisément le bitrate, qui correspond au débit d'information par seconde, de la façon suivante : $\text{bitrate} = n \times \text{échantillonnage}$. Ce dont nous disposons donc pour analyser un signal, est la donnée de l'amplitude en fonction du temps la caractérisant.

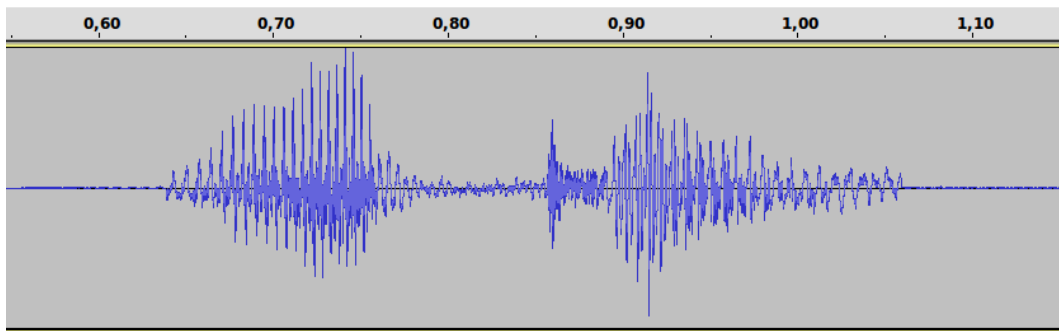


FIGURE 1.1 – Exemple audiogramme prononciation du mot "VICA"

1.2.3 Pré accentuation des aigus

Il a donc fallu pondérer l'influence des différentes bandes de fréquences pour rester le plus fidèle possible à la perception de l'oreille humaine. Ainsi, le poids des hautes fréquences (supérieures à 1000 Hz) a été diminué tandis que le poids des basses fréquences (inférieur à 1000 Hz) a été augmenté. Pour cela, on affecte l'amplitude d'une fréquence f à une nouvelle fréquence $mel(f)$ selon la règle :

1.2.4 Etalonnage des fréquences

L'oreille humaine perçoit les fréquences de façon non linéaire. Ainsi, l'augmentation d'une octave correspond en réalité à la multiplication de la fréquence par 2 : par exemple, le la 3 correspond à une fréquence de 440 Hz tandis que le la 4 correspond à une fréquence de 880 Hz.

1.3 Enregistrement, recadrage, filtrage HF

1.3.1 Synchronisation

Afin de synchroniser le début des enregistrements d'un mot, et de leur donner la même durée, il a été nécessaire 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 finit 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. Cependant, pour l'enregistrement de notre base de données, une pièce calme et un micro de bonne qualité nous ont permis un découpage satisfaisant. Cependant, l'utilisateur ne pouvant pas toujours se placer dans ces conditions, le signal est traité par un filtre anti-bruit.

Ce filtre consiste en l'utilisation de bibliothèques, SoX et ffmpeg, 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 unes des autres.

Ceci a été fait de manière empirique sur plusieurs enregistrements de mots différents, permettant une découpe automatique la plu satisfaisante possible pour l'ensemble des mots.

1.4 Echantillonnage, fenêtrage

Nam eu sollicitudin massa. Duis sagittis velit mi. Nunc dictum risus ac interdum lacinia[1]. Aliquam a fermentum lectus. Praesent dapibus molestie mauris sed vestibulum. Curabitur sodales egestas est a pellentesque. Quisque id vulputate erat, a sagittis turpis. Proin vulputate congue ante, a laoreet orci posuere venenatis. Suspendisse ullamcorper ac dolor nec dapibus.

1.4.1 début du filtre 2

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas rutrum aliquet odio ac volutpat. Suspendisse massa enim, faucibus ut mi dignissim, fringilla ultrices dolor. Quisque bibendum vel nisl eget interdum. Phasellus sed ultricies orci. Aenean egestas risus ut ante porttitor tempus. Suspendisse malesuada, neque ut ullamcorper hendrerit, nibh justo porta purus, quis suscipit neque nisi in nunc. Pellentesque est nisi, porta semper mauris quis, commodo ultricies turpis.

1.4.2 fin du filtre 2

Vivamus posuere malesuada metus, et fringilla felis congue sed. Fusce a nibh in urna molestie viverra vel in est. Donec ipsum nibh, commodo tristique bibendum vitae, consequat quis ipsum. Donec sed diam aliquet velit aliquam viverra. Etiam tincidunt egestas nibh, sit amet facilisis quam vulputate scelerisque. Maecenas nec eros rutrum, tempus ipsum sit amet, tempus dui. Quisque adipiscing varius neque, at egestas erat consequat vitae. Donec sit amet urna iaculis, pharetra arcu in, porttitor quam. Ut nunc ante, tincidunt placerat libero eu, hendrerit lacinia leo. Nam sapien enim, feugiat nec orci eget, porta suscipit diam. Nulla mattis molestie est vitae blandit. Nullam bibendum tempus odio, quis vestibulum quam consectetur at.

1.5 Transformée de Fourier

Nam eu sollicitudin massa. Duis sagittis velit mi. Nunc dictum risus ac interdum lacinia[1]. Aliquam a fermentum lectus. Praesent dapibus molestie mauris sed vestibulum. Curabitur sodales egestas est a pellentesque. Quisque id vulputate erat, a sagittis turpis. Proin vulputate congue ante, a laoreet orci posuere venenatis. Suspendisse ullamcorper ac dolor nec dapibus.

1.6 MEL

Nam eu sollicitudin massa. Duis sagittis velit mi. Nunc dictum risus ac interdum lacinia[1]. Aliquam a fermentum lectus. Praesent dapibus molestie mauris sed vestibulum. Curabitur sodales egestas est a pellentesque. Quisque id vulputate erat, a sagittis turpis. Proin vulputate congue ante, a laoreet orci posuere venenatis. Suspendisse ullamcorper ac dolor nec dapibus.

1.7 Transformée inverse

Nam eu sollicitudin massa. Duis sagittis velit mi. Nunc dictum risus ac interdum lacinia[1].

2. Les MMC

2.1 Principe, modèle discret

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas dapibus scelerisque elit pulvinar ornare. In sollicitudin dui eu eros feugiat, sit amet eleifend ante dapibus. Donec id aliquet dolor, at ullamcorper nibh. Nulla egestas odio tempus nunc venenatis fringilla. Fusce in risus eu augue ullamcorper euismod nec et lacus. Nullam elementum adipiscing tellus, vel aliquet tortor ultricies quis. In justo libero, dictum sed risus sodales, hendrerit tempor eros. Sed varius metus velit, nec placerat neque tincidunt ac.

2.2 Phase d'apprentissage

Morbi cursus risus eget scelerisque interdum. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam eget nibh tellus. In dolor purus, viverra nec eleifend id, commodo dignissim ipsum. Nullam viverra turpis at sodales commodo. Nunc sollicitudin nisi sit amet purus gravida, a porttitor velit bibendum. Cras varius nibh purus, vel sollicitudin lorem adipiscing id.

2.3 Phase de reconnaissance

Fusce commodo bibendum malesuada. Suspendisse ultrices pulvinar nulla, at ornare mauris accumsan tempus. Donec a mi libero. Fusce ut ante massa. Sed pretium augue ante, eu pellentesque nibh hendrerit sollicitudin. Maecenas placerat vehicula nulla, id tincidunt lectus placerat a. Vivamus et mattis quam, in sagittis felis. Aenean lobortis cursus nibh, eget vulputate mauris porta vel. Praesent at diam dictum, varius arcu id, sollicitudin lacus. Nullam sit amet mi non diam rutrum tincidunt. Pellentesque pretium magna velit, ac sagittis lectus iaculis non. Aenean vulputate fermentum tellus, non volutpat neque bibendum quis. Praesent malesuada orci quis diam ullamcorper iaculis. Nullam vitae purus euismod, fermentum mi a, mattis purus.

Deuxième partie

Approche commerciale

1. Côté web

Cras accumsan, sapien a mattis cursus, augue turpis rhoncus leo, et mattis lorem mi sed nisi. Fusce quam tortor, vulputate et tincidunt in, condimentum quis risus. Phasellus facilisis est id ipsum fermentum, eget aliquet arcu tristique. Mauris lacinia urna non risus ullamcorper auctor. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aenean a tincidunt mauris. Maecenas felis massa, tempus nec lorem et, accumsan iaculis nunc.

Suspendisse tincidunt eros nibh, mollis feugiat elit feugiat eget. Ut porta gravida orci in congue. Vestibulum et lectus adipiscing purus tempor feugiat. Morbi consectetur eget tellus ut feugiat. Sed ipsum nisl, feugiat vitae volutpat et, accumsan eget orci. Proin non leo eros. Vivamus malesuada neque et erat bibendum malesuada.

In hac habitasse platea dictumst. Vivamus eleifend erat ac quam aliquam, vehicula ultricies arcu iaculis. Suspendisse potenti. Quisque at commodo quam. Etiam ultricies elit leo, vel congue lorem tristique eget. Quisque sodales dignissim diam, ac malesuada libero scelerisque sed. Aliquam bibendum luctus elit vel suscipit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Quisque quis mi at mi vulputate rhoncus.

Integer convallis id libero vel egestas. Cras ultricies porta nisl, ut tempus purus pulvinar tristique. Duis ut hendrerit libero. Aliquam sit amet erat sit amet dolor sagittis feugiat. Nulla elit enim, iaculis et congue eget, porta hendrerit purus. Aliquam in facilisis metus. Vivamus euismod adipiscing pellentesque. Fusce a pellentesque lorem. Sed pretium ut ligula non blandit. Phasellus vitae mollis dolor. In a eleifend magna, non mollis magna. Etiam nisi tellus, blandit ut ante ut, vehicula tristique nunc.

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'elle se réalise plus tard que prévu. Le principal obstacle à l'explosion de cette technologie étant le manque de fiabilité totale, 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. 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, notamment le célèbre Puma de l'armée française. 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, 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, 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, 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, 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

Praesent quis mauris tellus. Nulla ut dui vel ligula aliquet posuere. Morbi dignissim pellentesque consequat. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin eleifend feugiat enim, nec egestas neque semper non. Vestibulum ullamcorper lorem eu dui pretium, eu pretium odio pellentesque. Mauris sed fringilla nisl. Sed ut vestibulum elit. Aenean lobortis rutrum ligula sit amet convallis. Praesent fermentum malesuada felis at interdum. Donec iaculis nunc lectus, ut facilisis turpis accumsan sit amet. Nunc elementum ut lacus sed elementum. Morbi rutrum tristique dignissim.

Sed interdum augue ac velit tristique, eget lacinia mauris consectetur. Fusce aliquet arcu ut lectus congue volutpat. Sed lobortis leo sit amet neque gravida egestas. Praesent diam nulla, adipiscing eu ullamcorper et, posuere ac lectus. Nulla adipiscing diam in felis cursus, non eleifend sapien vehicula. Maecenas et magna metus. Nunc fringilla risus sapien, nec sagittis arcu adipiscing quis. Cras vitae consequat leo, sagittis bibendum lorem. Quisque interdum, nibh a interdum blandit, augue nisl fringilla neque, sed sollicitudin tortor leo id lacus. Ut vel euismod eros, sed bibendum quam. Etiam interdum elit ut orci blandit, vitae vulputate nunc gravida. Aenean ipsum tellus, facilisis eget ipsum id, posuere interdum enim. Suspendisse vitae leo egestas, tempus erat sit amet, malesuada tortor.

Quisque a tempus metus. Fusce ornare accumsan libero porttitor feugiat. Nullam tempor velit vel cursus commodo. Fusce condimentum augue felis, sed varius quam varius vel. Etiam et neque vitae ante ultricies mattis. Suspendisse vehicula ligula felis, at vestibulum nisl dictum et. Morbi ante odio, lacinia et massa at, posuere sollicitudin ante. Fusce dignissim laoreet sagittis. Aenean non egestas enim. Vestibulum rutrum, nisl vel condimentum fringilla, tellus justo adipiscing est, ut sollicitudin mauris nisi dapibus felis. Aliquam diam elit, pretium rhoncus malesuada mattis, tincidunt non ipsum. Duis ut ipsum bibendum libero porttitor placerat vel et velit.

Troisième partie

Le Code

1. Python

```
1 # -*- coding: utf-8 -*-
2
3 import numpy as np
4 import scipy as sc
5 import math
6 from operator import add
7
8
9 TAILLE_TABLEAU_MEL_ENTREE = 24
10
11 NOMBRE_COMPOSANTES_GARDEES = 13
12
13 B = TAILLE_TABLEAU_MEL_ENTREE
14
15
16 def inverseDCTI(x): # x represente le tableau en mel donne par les
    fonctions precedentes
17     X = np.zeros(B)
18     for k in range(B):
19         X[k] = (0.5 * (x[0] + math.pow(-1, k) * x[B-1]) + \
20             reduce(add, [x[n] * math.cos(math.pi * n * k / (B -
21                 1)) \
22                     for n in range(1,B - 1)])) * math.sqrt(2. / (B -
23                 1))
24     return X
25
26 def inverseDCTII(x):
27     X = np.zeros(B)
28     for k in range(B):
29         X[k] = reduce(add, [x[n]*math.cos(math.pi*(n + 0.5)
30             * k / B) \
31                 for n in range(B)]) * math.sqrt(2. / B)
32     return X
33
34 def inverseDCTIII(x):
35     X = np.zeros(B)
36     for k in range(B):
```

```

34         X[k] = (0.5*x[0]+reduce(add, [x[n] *
math.cos(math.pi * (k + 0.5) / B)\
35         for n in range(1, B)])) * math.sqrt(2. / B)
36     return X

```

2. C/C++

```
1 cDouble* fftCT(cDouble *sig)
2 {
3     int N = sizeof(sig)/sizeof(cDouble);
4     //int iMax = (int)(log(N)/log(2));
5     int i,j,k,p=0,f=1;
6     cDouble ekN;
7     cDouble **tmp = (cDouble**) malloc(2*sizeof(cDouble*));
8
9     for (i=0;i<2;i++)
10         tmp[i] = (cDouble*) malloc(N*sizeof(cDouble));
11     for (i=0;i<N;i++)
12         tmp[0][i] = sig[i];
13     for (i=N/2;i!=1;i/=2)
14     {
15         for (j=0;j<i;j++)
16             for (k=0;k<N/(2*i);k++)
17             {
18                 ekN = e(k,N)*tmp[p][i*(2*k+1)+j];
19                 tmp[f][i*k+j] = tmp[p][i*(2*k)+j] + ekN;
20                 tmp[f][i*k+j+N/2] = tmp[p][i*(2*k)+j] - ekN;
21             }
22         p = f;
23         f = (p+1)%2;
24     }
25     return tmp[p];
26 }
```

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 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 1000euros) 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 5euros. Il est difficile de prévoir le potentiel de cette version, puisque les concurrents sont très nombreux, de qualité et de prix très variables.

Bibliographie

- [1] Lawrence Rabiner. *Fundamentals of Speech Recognition*. Prentice Hall PTR, 1993.