

SpeechApp

Un logiciel de reconnaissance
automatique de la parole



Plan

1. Introduction
2. Fonctionnement de la reconnaissance vocale
3. Etude de marché
4. Développement d'un produit commercialisable
5. Etude économique
6. Conclusion



Le fonctionnement de la reconnaissance vocale



Qu'est-ce que le son ?

- Onde mécanique de pression
- Amplitude et fréquences caractéristiques
- Vitesse de 340 m.s^{-1}



L'enregistrement

Deux approches d'enregistrement du son :

- Tympan : mesure la fréquence de vibration de l'onde
- Micro : mesure l'amplitude de l'onde à intervalles de temps réguliers.
- Fréquence d'échantillonnage : 44100 Hz



Gestion du bruit

- Le bruit : une perturbation du son dû à l'environnement (voix, sons urbains...)
- Le bruit a des fréquences plutôt basses
- Passe-haut : filtre réduisant l'amplitude des basses fréquences



Sofiane



Objectifs de ce qui va suivre

- Nous avons réussi à obtenir le spectre correspondant à un certain instant de l'échantillon ; il faut maintenant :



→ Générer des motifs reconnaissables que l'on utilisera lors de la comparaison à partir de ce spectre

- Ces motifs doivent répondre à plusieurs exigences :

→ Conserver le maximum d'information pertinente sur le signal...

→ Eviter la redondance d'information

→ Etre légers...

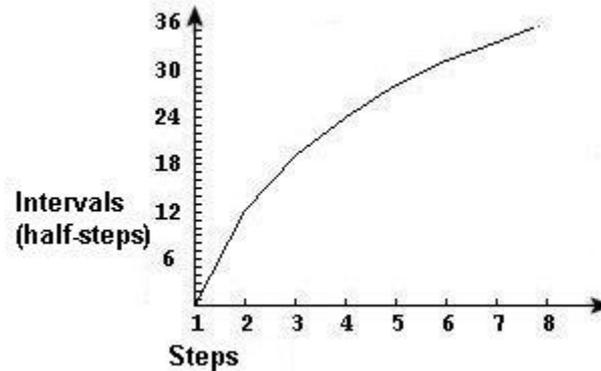
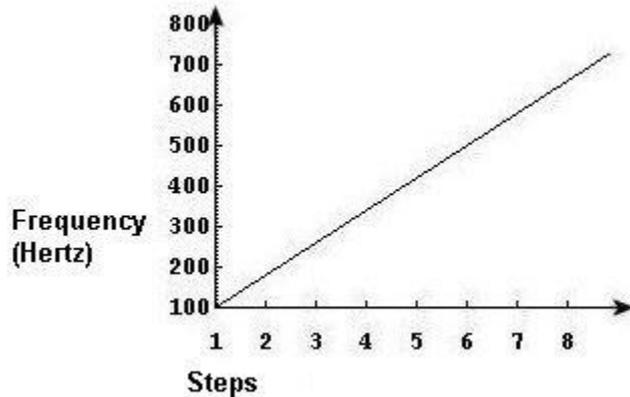


MEL : UNE ÉCHELLE PLUS PROCHE DE L'OREILLE

- L'oreille humaine ne perçoit pas les fréquences selon une échelle linéaire mais selon une échelle LOGARITHMIQUE.
- Les mesures aux hautes et aux basses fréquences ne nécessitent pas la même précision.



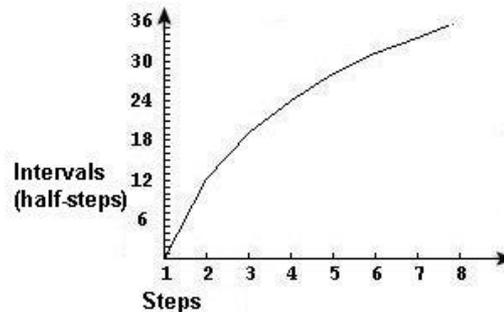
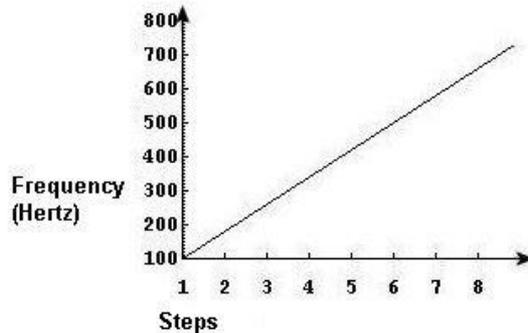
On cherche à exprimer les fréquences dans une nouvelle échelle permettant cette différenciation haute/basse fréquence.



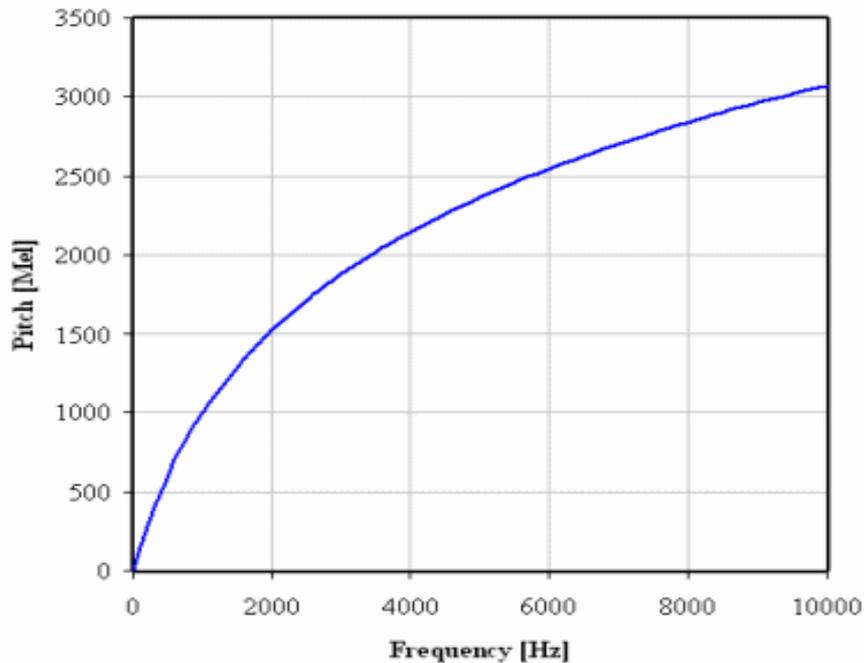
MEL : UNE ÉCHELLE PLUS PROCHE DE L'OREILLE

- L'oreille humaine ne perçoit pas les fréquences selon une échelle linéaire mais selon une échelle LOGARITHMIQUE.
- Les mesures aux hautes et aux basses fréquences ne nécessitent pas la même précision.

→ **On cherche à exprimer les fréquences dans une nouvelle échelle permettant cette différenciation haute/basse fréquence.**



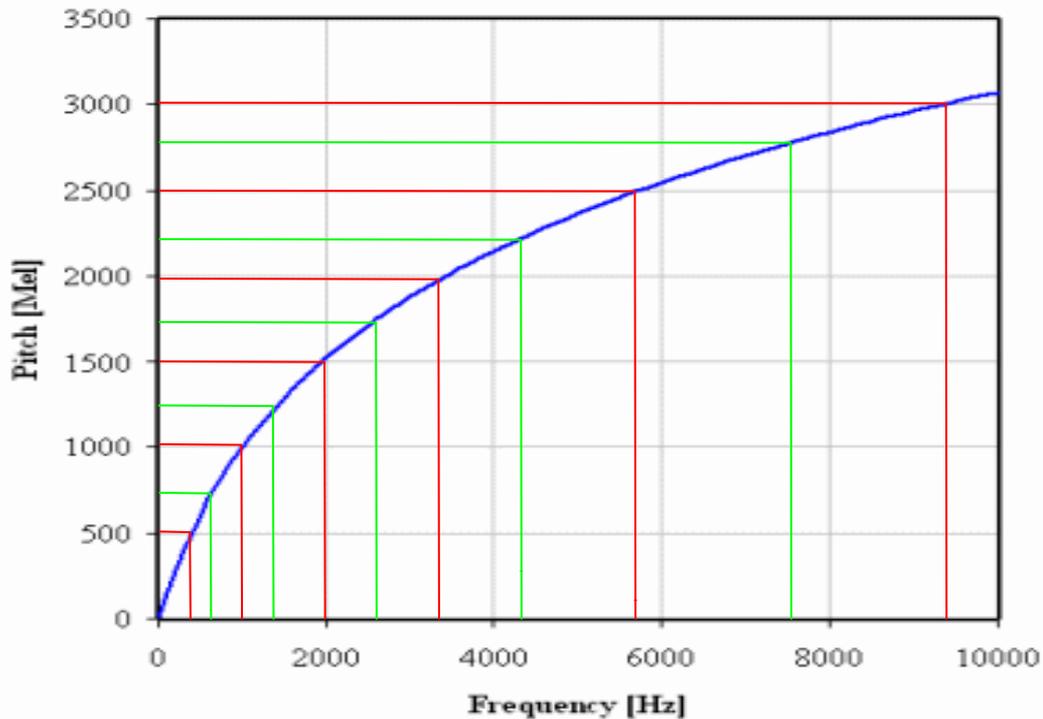
MEL : UNE ÉCHELLE PLUS PROCHE DE L'OREILLE



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



MEL : UNE ÉCHELLE PLUS PROCHE DE L'OREILLE



Bilan



Nouvelle représentation du signal obtenue à partir du spectre, plus en adéquation avec la façon dont l'oreille perçoit les fréquences.

Il s'agit d'un tableau de 24 cases

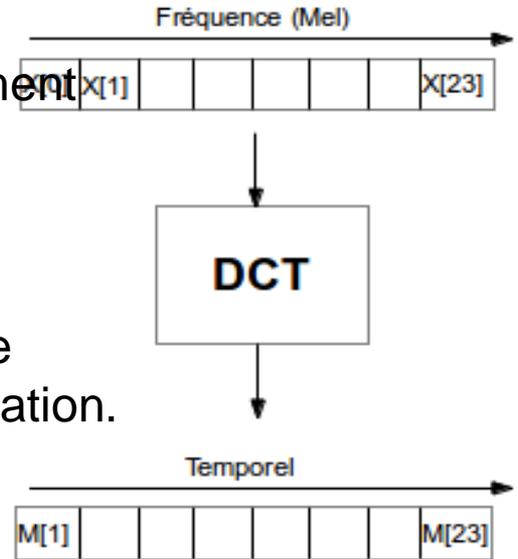
Ce tableau rempli presque tous les objectifs cités, cependant du progrès peut être fait pour supprimer la redondance d'information et faire encore l'économie de quelques cases...



On repasse en temporel !!

L'échelle Mel est une échelle de fréquence donc le tableau de 24 cases dont nous disposons actuellement est échellonné en fréquences.

Repasser à un échellonnement en temporel décorrelle certaines valeurs et évite ainsi la redondance d'information.



→ DECOMPOSITION EN COSINUS INVERSE

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



La récompense : les MFCC

Mel Frequency Cepstral Coefficient

Constitue l'empreinte du signal à un instant donné

Pour l'obtenir : 11 premières cases du tableau précédent +
12ème valeur : énergie du signal à l'instant considéré

→ On est prêt à passer à la reconnaissance ...

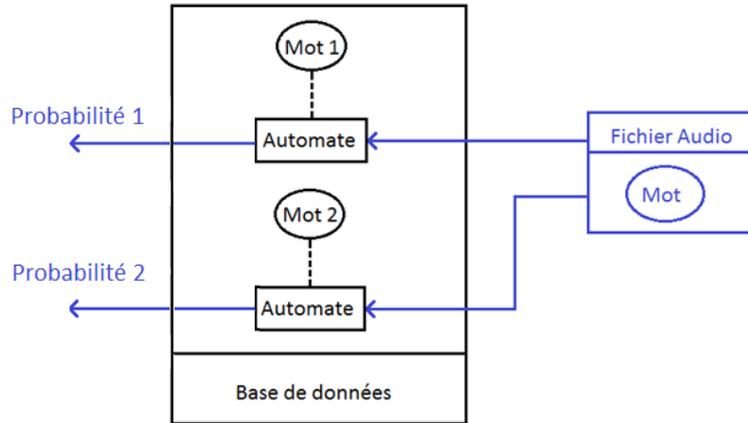


La reconnaissance



Comment reconnaître le mot ?

On « cherche » le mot prononcé parmi la base de mots reconnus



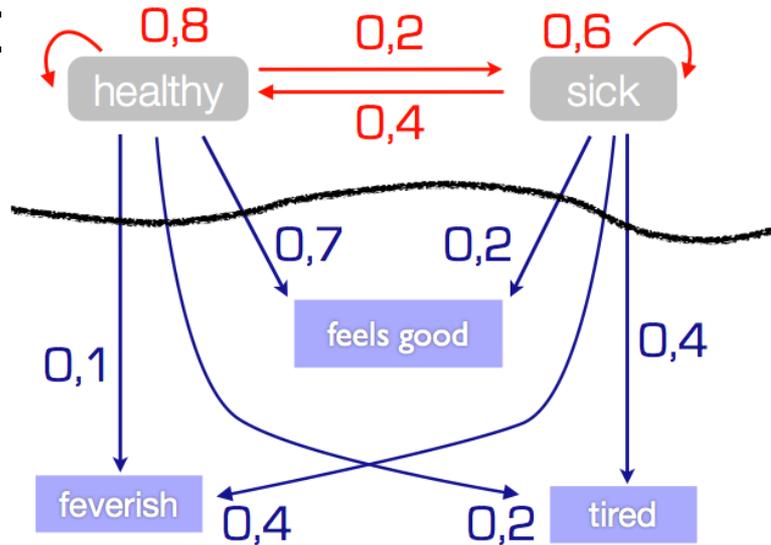
On obtient, pour chaque mot de la base, la probabilité que ce soit le bon mot

Le mot prononcé est alors *a priori* le mot qui a obtenu la plus grande probabilité



Markov à la rescousse

Chaque modèle de Markov va servir à stocker un mc



Chaque état correspond à un phonème

Chaque sortie correspond à un son

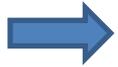
Les transitions sont associées à des probabilités



Forward !

Utilisation : l'algorithme qui calcule les probabilités

L'algorithme parcourt tous les chemins et regarde s'il correspond au mot prononcé



On obtient une probabilité que
enregistrement=mot



Création des modèles

Plusieurs enregistrements du mot par locuteur

Après traitement du son, on découpe chaque enregistrement en phonèmes Donne
les états

Reste à faire le reste de l'automate



Le reste de l'automate

Au début, on met la même probabilité partout

Ensuite, on se sert des enregistrements pour **améliorer** l'automate

Algorithme de Baum-Welch



Baum-Welch

On procède par répétition :

- On regarde quelles transitions et quelles sorties servent pour tous les enregistrements
- On augmente les probabilités associées

On répète cette opération jusqu'à obtenir une bonne reconnaissance du mot



Quelques chiffres

Ordre de grandeur :

- Automate : taille, nombre d'états, probabilités
- Enregistrements et locuteurs
- Forward et Baum-Welch : temps et nombre d'itérations



Intégration globale de l'algorithme

- Architecture générale en Python
- Flux de données entre chaque « brique » de l'algorithme adaptés en entrée/sortie



Optimisation

- Certaines parties du programmes sont beaucoup plus gourmandes en ressources
- Utilisation du C pour augmenter grandement la vitesse (passerelle avec Python : Boost)



ETUDE DE MARCHÉ





SpeechAp

30 Janvier 2014

Etude de marché : applications de la reconnaissance vocale



Une technologie promise à un futur radieux... mais encore perfectible !



SpeechAp

30 Janvier 2014

Etude de marché : applications de la reconnaissance vocale



De nombreuses applications dans différents domaines professionnels :
médecine, droit, police, contrôle aérien...



SpeechAp

30 Janvier 2014

Etude de marché : applications de la reconnaissance vocale



Des nombreuses applications personnelles et ludiques



SpeechAp

30 Janvier 2014

Etude de marché : concurrence



- Un concurrent qui monopolise le marché professionnel
 - un marché « personnel/ludique » plus ouvert

ETUDE DE MARCHÉ



Développement d'un produit commercialisable



Notre Start Up doit faire des bénéfices !



Notre Start Up doit faire des bénéfices !

L'utilisateur doit cliquer sur ce bouton :



Développement d'un produit commercialisable

Nous disposons des algorithmes fonctionnels

Nous disposons des algorithmes fonctionnels

MAIS ils ne sont pas commercialisables tel quel !

Nous disposons des algorithmes fonctionnels

MAIS ils ne sont pas commercialisables tel quel !

Il faut créer des services payants utilisant ces algorithmes

Premier service : application de reconnaissance utilisant une base de donnée préconçue

Pour un client souhaitant effectuer la reconnaissance d'un mot

Rémunération : contenu publicitaire sur cette page

Deuxième service : possibilité de créer ses propres bases de données

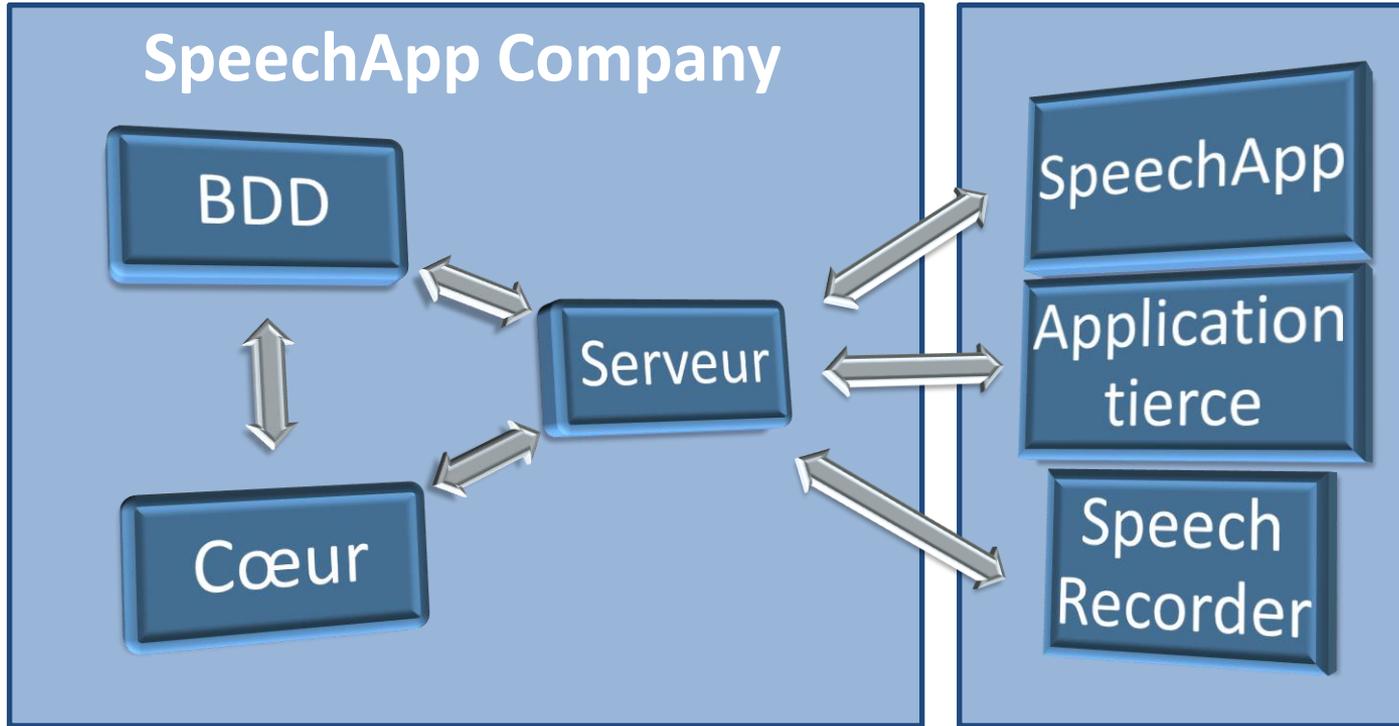
Pour un client voulant créer une application / un site utilisant SpeechApp

Rémunération : vente de licences

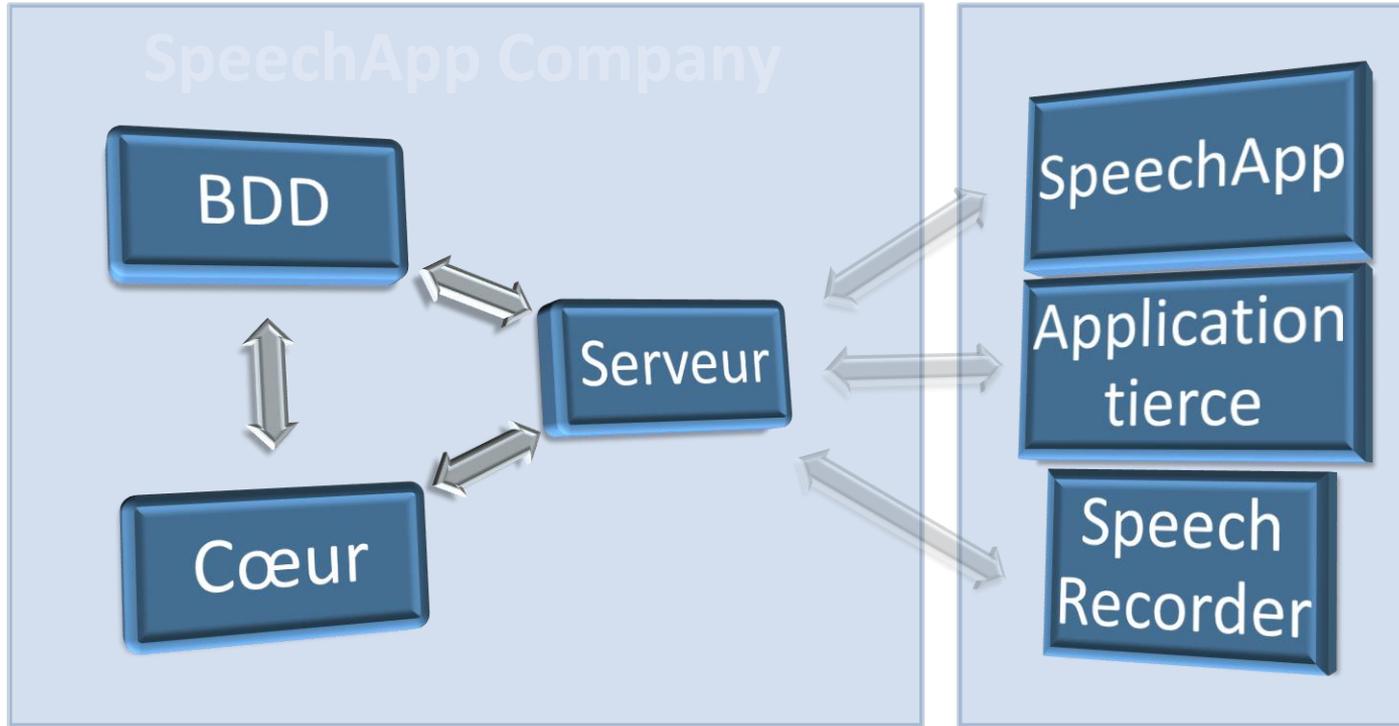
Ce dont on a besoin :

- Protection des algorithmes
- Protection des bases de données
- Maintenance facilitée

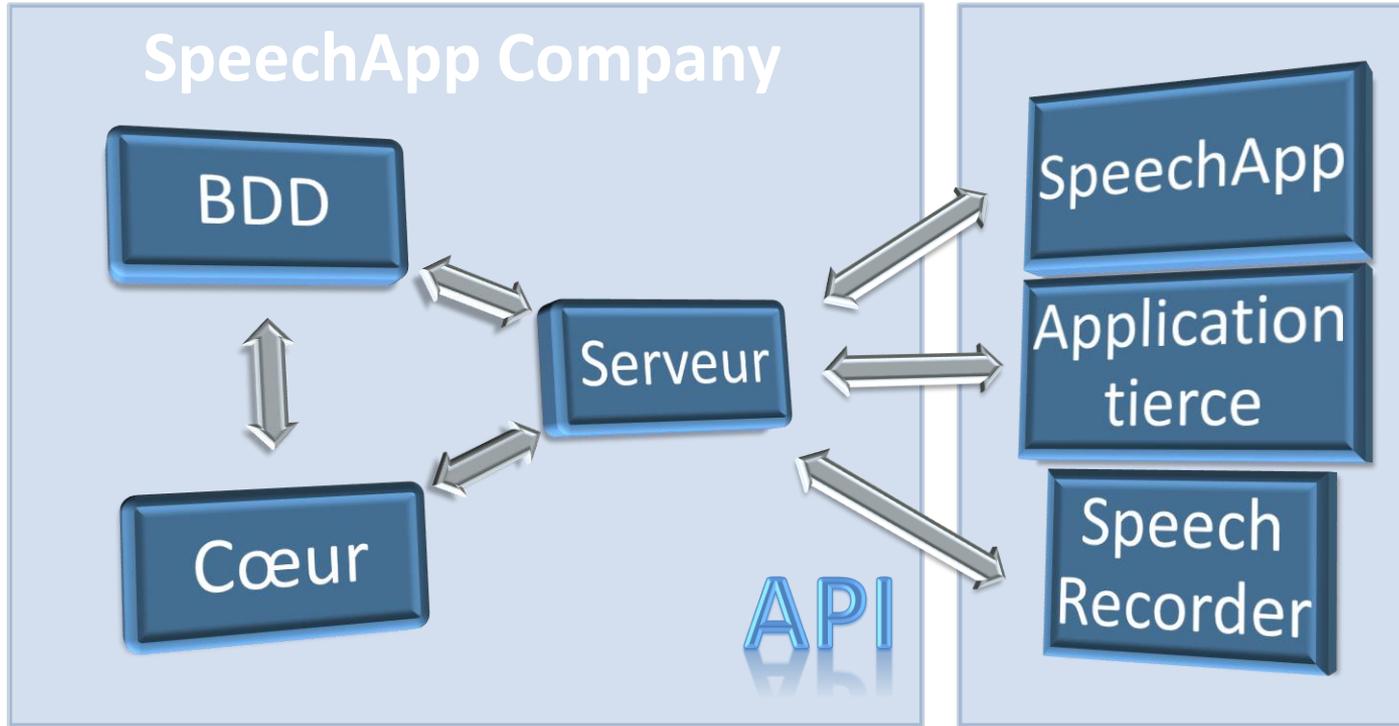
Structure Client/Serveur



Structure Client/Serveur

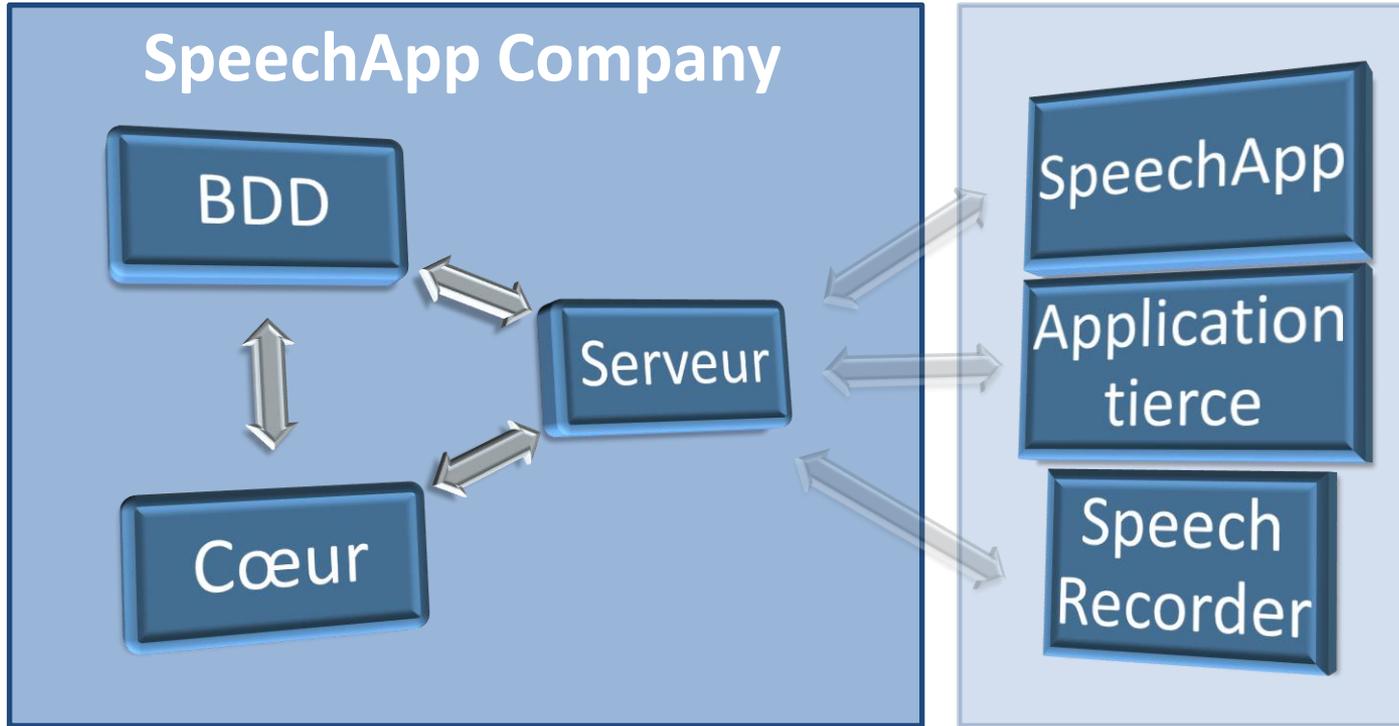


Structure Client/Serveur

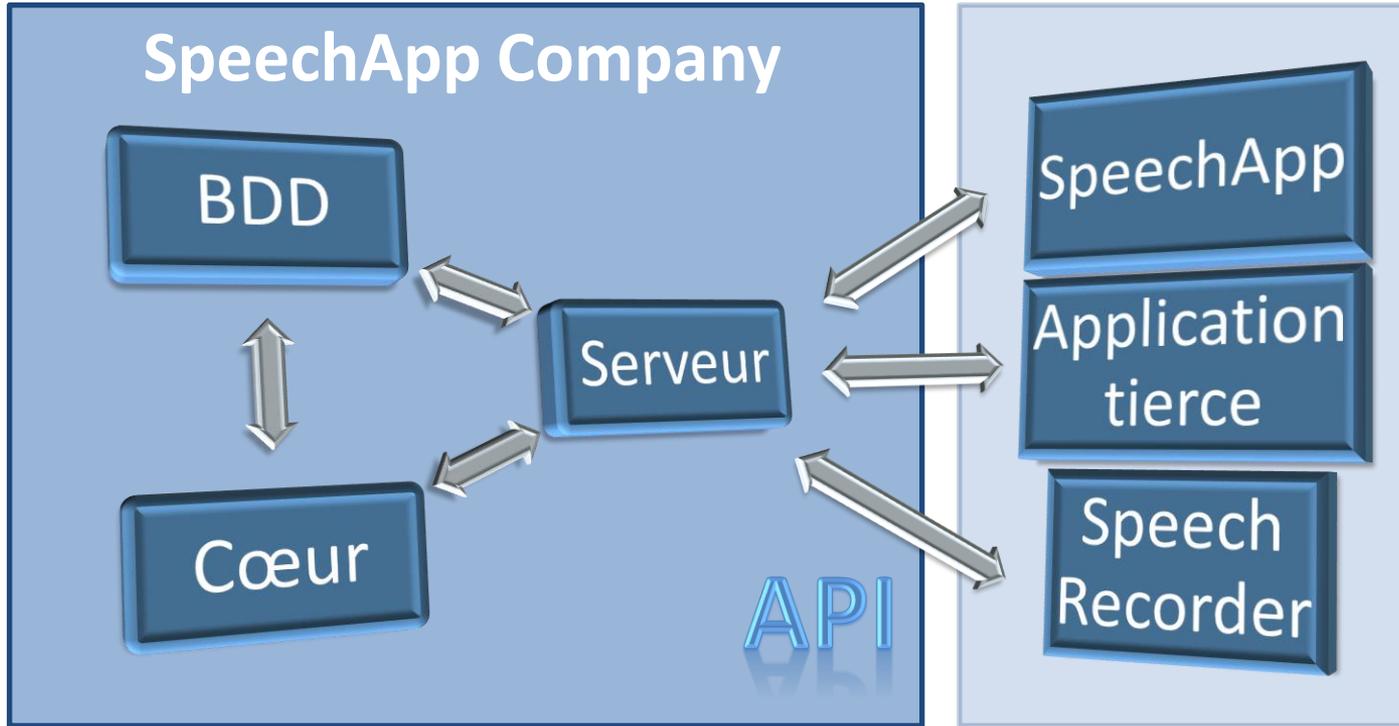


Application Programming Interface

Structure Client/Serveur

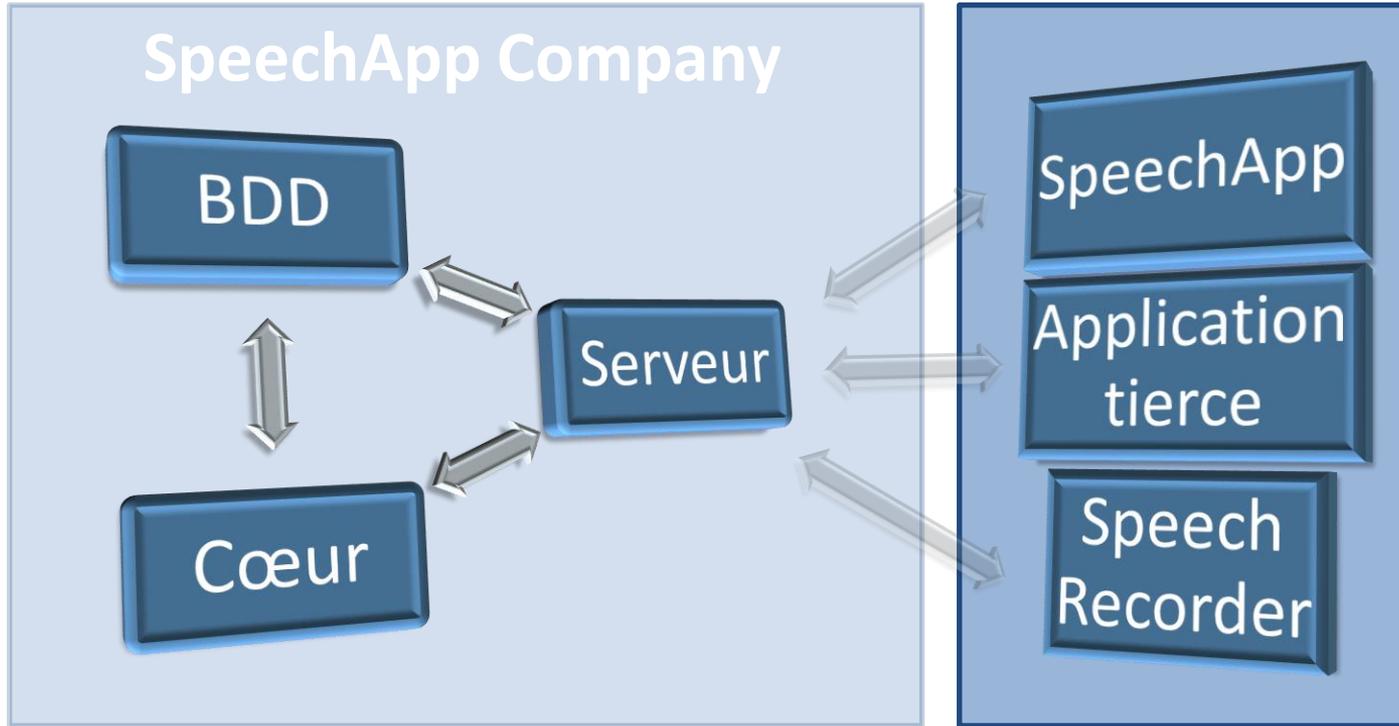


Structure Client/Serveur

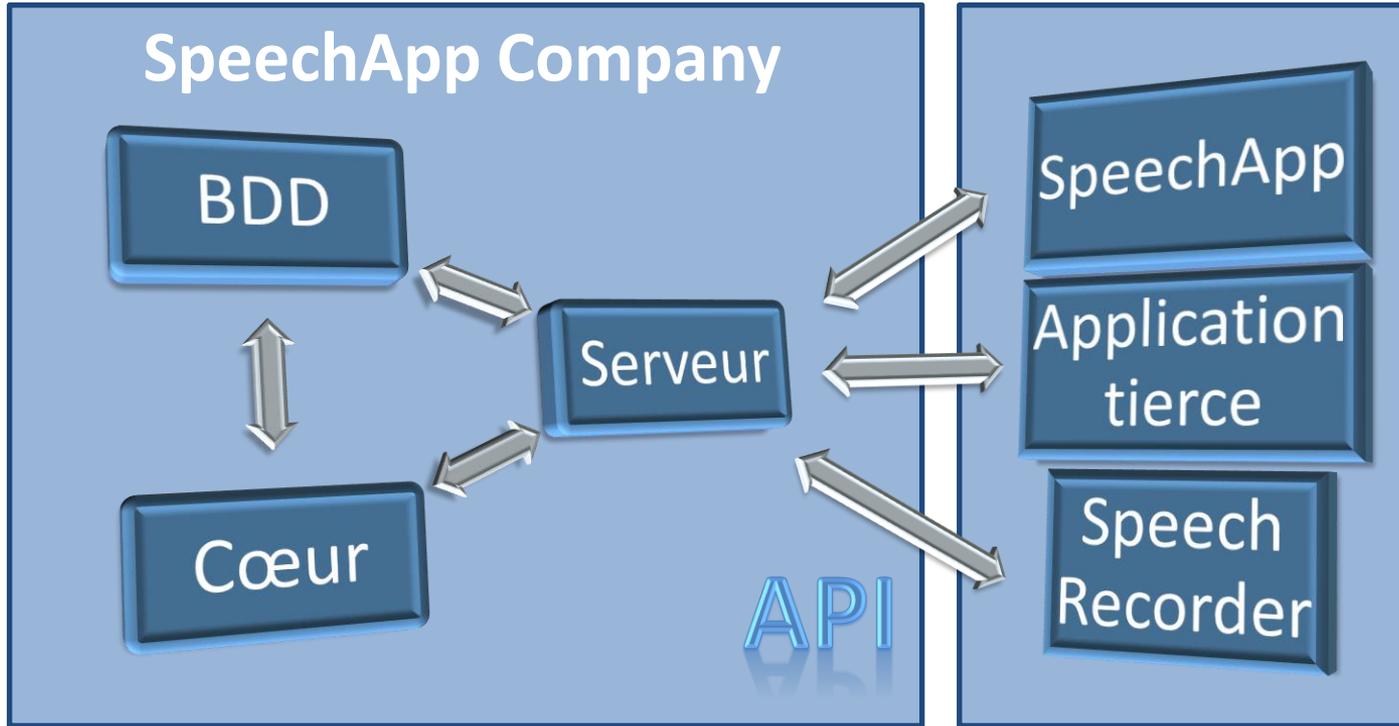


Application Programming Interface

Structure Client/Serveur



Structure Client/Serveur



Application Programming Interface

SPEECHAPP
ET
SPEECHRECORDER



Démonstrateurs

Démonstrateurs

Pour nos tests:

Démonstrateurs

Pour nos tests:

Pour permettre l'enregistrement simple de nouveaux mots

Démonstrateurs

Pour nos tests:

Pour permettre l'enregistrement simple de nouveaux mots

SpeechRecorder

Démonstrateurs

Pour nos tests:

Pour permettre l'enregistrement simple de nouveaux mots

SpeechRecorder

Commercialement:

Démonstrateurs

Pour nos tests:

Pour permettre l'enregistrement simple de nouveaux mots

SpeechRecorder

Commercialement:

Pour montrer les capacités de notre technologie

Démonstrateurs



Pour nos tests:

Pour permettre l'enregistrement simple de nouveaux mots

SpeechRecorder



Commercialement:

Pour montrer les capacités de notre technologie

SpeechApp

Démonstrateurs



Pour nos tests:

Pour permettre l'enregistrement simple de nouveaux mots

SpeechRecorder



Commercialement:

Pour montrer les capacités de notre technologie

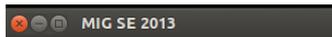
SpeechApp

SpeechRecorder

Une interface pour la consultation et la modification de plusieurs bases de données de mots.

SpeechRecorder

Une interface pour la consultation et la modification de plusieurs bases de données de mots.



Se connecter

Identifiant :

Mot de passe :

Se connecter

S'inscrire

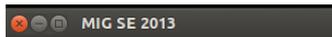
Identifiant :

Mot de passe :

S'inscrire

SpeechRecorder

Une interface pour la consultation et la modification de plusieurs bases de données de mots.



Se connecter

Identifiant :

Mot de passe :

Se connecter

S'inscrire

Identifiant :

Mot de passe :

S'inscrire

MIG SE 2013

Reconnaissance vocale

Se déconnecter

Quitter



Enregistrer

SpeechRecorder

Interface principale d'enregistrement de nouveaux mots dans des bases de données

MIG SE 2013

Reconnaissance vocale

Entrez le mot que vous souhaitez enregistrer

Avant de procéder aux enregistrements nécessaires, il convient d'enregistrer du bruit pour permettre efficacement le traitement du signal

Enregistrer du bruit

Il est nécessaire pour créer un modèle de Markov caché de procéder à une dizaine d'enregistrements du même mot
Vous aurez deux secondes à chaque enregistrement pour prononcer votre mot une fois l'acquisition lancée

Terminer l'enregistrement du HMM 3

Quitter

SpeechApp

Démonstrateur commercial du projet.

SpeechApp

Démonstrateur commercial du projet.

Reconnaissance vocale des 10 premiers entiers.

SpeechApp

Démonstrateur commercial du projet.

Reconnaissance vocale des 10 premiers entiers.

Basé sur les technologies ouvertes du Web.



SpeechApp

SpeechApp

Home

Our services

About ▾

SpeechApp !

Welcome to this demonstrator of our state of the art speech recognition technology

Test it, feed us with a few words

[Learn more »](#)

Demonstrator



SpeechApp

Fonctionne sur le modèle client-serveur décrit précédemment:
pas d'installation nécessaire !

SpeechApp

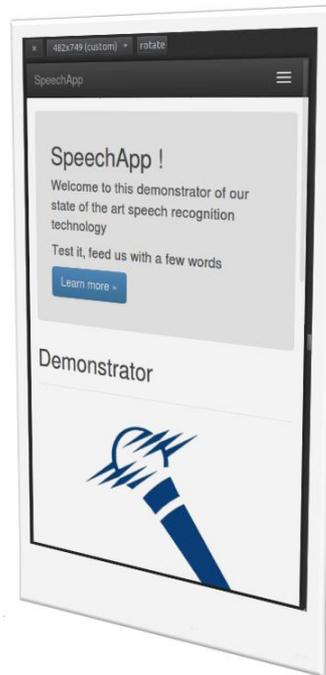
Fonctionne sur le modèle client-serveur décrit précédemment:
pas d'installation nécessaire !



<http://speechapp.tk/>

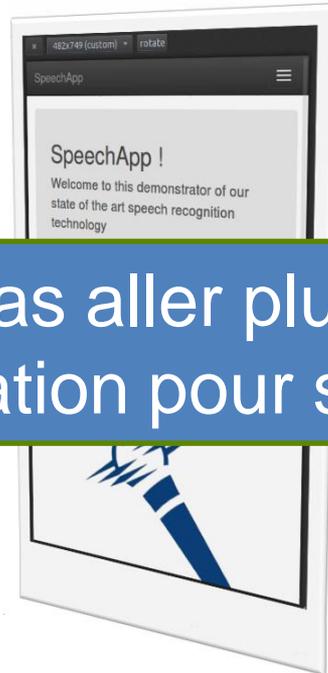
SpeechApp

S'adapte automatiquement à la taille de l'écran



SpeechApp

S'adapte automatiquement à la taille de l'écran



Pourquoi ne pas aller plus loin et montrer une application pour smartphone ?

SpeechApp

Nous l'avons fait !



SpeechApp

Nous l'avons fait !



<http://speechapp.tk/android>



SpeechApp

Nous l'avons fait !



<http://speechapp.tk/android>



Essayez SpeechApp pour Android

Développement commercial

Budget & modèle économique

- Stratégie envisagée :
 - Développement d'un site web
 - Mise en vente d'un logiciel pour particuliers sur ce site
 - Mise en vente de licences pour les professionnels

Salaires

- 13 employés sur le projet durant 1 mois :
 - 11 développeurs de moins de deux ans d'expérience (2250 € brut)
 - 2 membres des ressources humaines (2750 € brut)
- 22 % de charges sociales et 44 % de charges patronales
- Budget total : 44200 €

Résultats prévisionnels

COMPTE DE RÉSULTAT PRÉVISIONNEL

Produit				Charges	
Vente	Prix unité TTC	Nombre	Total	Salaires	44 121,60 €
Logiciel	5,00 €	26000	130 000,00 €	Frais pub (google)	26 000,00 €
Licences	1 000,00 €	10	10 000,00 €	2 serveurs	7 272,00 €
62606,40 €					

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	62710,40 €
Créances clients	0,00 €		
Trésorerie	0,00 €		

Bilan

Impôts		
Sur les sociétés	33% des bénéfices	11 447,57 €
TVA	20% sur les ventes	28 020,80 €

Bénéfices

34 689,60 €