

# Interrogation de transcription textuelle d'enregistrement audio de réunion

Livrable Unique - 12 mai 2012

Polytech Nantes – département Informatique

Binôme étudiant :

- Doriane PERESSE
- Maxime CATOIRE

Tuteurs enseignant :

- Marc GELGON
- Hoel LE CAPITAINE

Coordinateur :

- Jean-Pierre GUÉDON

Organisme commanditaire : Project2Cloud

Tuteur Industriel :

- Jean-Michel MAILLET



# Table des matières

Introduction	3
I Contexte du projet	4
II Cahier des charges	6
III Étude bibliographique	27
IV Procédure de recette	38
V Estimation des charges et planifications	40
VI Modélisation	41
VII Réalisation	47
VIII Tests	49
IX Manuel d'utilisation	51
X Dossier de déploiement	53
XI PV de recette	55
Conclusion	57
Références	58

# Introduction

Première partie

Contexte du projet

## 1.1 L'entreprise : Project2Cloud

Project2Cloud est une plate-forme de gestion collaborative. Une site de présentation est déjà en ligne à l'adresse <http://project2cloud.org/>. Elle propose différents services de communication, tels que la vidéo-conférence, le chat et la conversation audio. Le but est d'offrir aux utilisateurs un panel complet de communication, leur permettant d'échanger en temps réel des informations et des services. Project2Cloud innove également avec la possibilité d'associer aux sessions de dialogue des documents de différents types (texte, vidéos...). Le document associé à la session permet de créer une "mémoire" numérique du document, autour de laquelle de nouvelles connaissances peuvent venir s'ajouter à tout moment.

P2C (abréviation courante de Project2Cloud) est géré par trois informaticiens passionnés depuis maintenant 1 an. La plate-forme P2C est pour le moment en bêta-test privé, c'est à dire qu'elle n'est pas encore accessible. Elle est actuellement testée par plusieurs industriels. Pour les besoins du projet, un accès nous a été donné à la version opérationnelle de la plate-forme. Cette version nous permet d'avoir une idée de l'application qui sera prochainement disponible.

## 1.2 Notre projet : Transcriptions audio vers texte pour des réunions enregistrées

Notre projet s'implantera directement dans la plate-forme P2C. Comme nous l'avons expliqué précédemment, P2C permettra aux utilisateurs de communiquer facilement grâce à un panel complet d'échanges vidéo, audio ou texte. Le but de notre projet est de faciliter l'indexation des conférences vidéos (c'est à dire de mieux séquencer temporellement les vidéos enregistrées). Pour cela, nous devons ajouter une fonctionnalité supplémentaire : la transcription vocale des conversations. Concrètement, lors d'une conférence vidéo, les conversations des différents participants seront traduites en texte. Ce même texte sera enregistré sur un serveur. Le but est de permettre à un utilisateur de retrouver un mot ou un groupe de mot prononcé pendant une conférence, et de pouvoir visionner cette dernière à l'instant où le mot a été prononcé. On crée ainsi une "mémoire" des conférences, avec la possibilité de retrouver des informations très précises à des moments clés. Notre projet est donc de convertir les conversations en texte, de les enregistrer, et de pouvoir indexer ces conversations à partir de mots clés.

Deuxième partie

Cahier des charges

# Table des matières

---

<b>1</b>	<b>Fondements du projet</b>	<b>10</b>
1.1	Contexte du projet . . . . .	10
1.2	Objectifs du projet . . . . .	10
<b>2</b>	<b>Personnes et organismes impliqués dans le projet</b>	<b>11</b>
2.1	Maître d'ouvrage . . . . .	11
2.2	Tuteurs enseignants . . . . .	11
2.3	Elèves Étudiants . . . . .	11
2.4	Utilisateurs du produit . . . . .	12
<b>3</b>	<b>Contraintes sur le projet</b>	<b>13</b>
3.1	Contraintes négociables . . . . .	13
3.1.1	Types de modules externes utilisés . . . . .	13
3.1.2	Environnement de développement du projet . . . . .	13
3.1.3	Langage reconnu par les modules de transcription . . . . .	13
3.1.4	Fonctionnalités supplémentaires . . . . .	14
3.2	Contraintes non négociables . . . . .	14
3.2.1	Contraintes de temps . . . . .	14
3.2.2	Environnement de fonctionnement du projet . . . . .	14
3.2.3	Lieu de fonctionnement . . . . .	14
3.2.4	Contraintes matérielles . . . . .	14
3.2.5	Quel est le budget affecté au projet ? . . . . .	14
3.3	Faits et hypothèses utiles . . . . .	15
<b>4</b>	<b>Exigences fonctionnelles</b>	<b>16</b>
4.1	Portée du travail . . . . .	16
4.1.1	La situation actuelle . . . . .	16
4.1.2	Contexte du travail . . . . .	18
4.1.3	Portée du produit (cas d'utilisations) . . . . .	18



4.2	Exigences fonctionnelles et exigences sur les données . . . . .	19
4.2.1	Exigences fonctionnelles . . . . .	19
4.2.2	Exigences sur les données . . . . .	19
<b>5</b>	<b>Exigences non fonctionnelles</b>	<b>20</b>
5.0.3	L'interface . . . . .	20
5.0.4	Le style du produit (packaging inclus) . . . . .	20
5.1	Facilité d'utilisation et facteurs humains . . . . .	20
5.1.1	Facilité d'utilisation . . . . .	20
5.1.2	Personnalisation et internationalisation . . . . .	20
5.1.3	Exigences d'accessibilité . . . . .	21
5.2	Fonctionnement du produit . . . . .	21
5.2.1	Rapidité d'exécution et temps de latence . . . . .	21
5.2.2	Précision et exactitude . . . . .	21
5.2.3	Fiabilité et disponibilité . . . . .	21
5.2.4	Robustesse ou tolérance à un emploi erroné . . . . .	21
5.2.5	Capacité de stockage et montée en charge . . . . .	22
5.2.6	Adaptation du produit à une augmentation de volume à traiter . . . . .	22
5.2.7	Longévité . . . . .	22
5.3	Adéquation du produit avec son environnement . . . . .	22
5.3.1	Environnement physique prévu . . . . .	22
5.3.2	Environnement technologique prévu . . . . .	22
5.3.3	Applications « partenaires » (avec lesquelles le produit doit collaborer) . . . . .	22
5.3.4	Approche « produit » prêt à être commercialisé . . . . .	23
5.4	Maintenance, support, portabilité, installation du produit . . . . .	23
5.4.1	Maintenance du produit . . . . .	23
5.4.2	Conditions spéciales concernant la maintenance du produit . . . . .	23
5.4.3	Installation du système . . . . .	23
5.5	Sécurité . . . . .	23
5.5.1	Accès au système . . . . .	23
5.5.2	Intégrité . . . . .	23
5.5.3	Protection des données à caractère personnel . . . . .	23
5.5.4	Audit et traçabilité . . . . .	23
5.5.5	Protection contre les infections . . . . .	24
5.6	Exigences culturelles et politiques . . . . .	24
5.6.1	Exigences culturelles . . . . .	24
5.6.2	Exigence politiques . . . . .	24
5.7	Lois et standards influençant le produit . . . . .	24
5.7.1	Conformité avec la loi . . . . .	24
5.7.2	Conformité avec des standards . . . . .	24

<b>6</b>	<b>Autres aspects du projet</b>	<b>25</b>
6.1	Questions sans réponse . . . . .	25
6.2	« COTS » : Progiciels ou composants commerciaux . . . . .	25
6.3	Nouveaux problèmes, créés par l'apparition du nouveau système	25
6.4	Tâches à faire pour livrer le système . . . . .	25
6.5	Contrôle final de qualité sur site (Cutover) . . . . .	25
6.6	Risques liés au projet . . . . .	26
6.7	Estimation des coûts du projet . . . . .	26
6.8	Manuel utilisateur et formations . . . . .	26
6.9	Salle d'attente : idées pour les futures versions . . . . .	26

---

# Chapitre 1

## Fondements du projet

### 1.1 Contexte du projet

Le projet proposé est le fruit de la collaboration de l'école polytechnique de l'université de Nantes et de l'entreprise Project2Cloud. Il sera mené de bout en bout par deux étudiants de quatrième année du cycle ingénieur, supervisés par deux professeurs référents. Il s'étendra de septembre à mai.

### 1.2 Objectifs du projet

L'objectif est de permettre aux étudiants en charge de réaliser un projet professionnel dans le cadre de leurs études. L'objectif du projet est d'ajouter à une plateforme de communication existante un module de transcription textuelle. Chacune des vidéos enregistrées sur la plateforme aura donc un fichier texte associé, qui correspondra à la transcription textuelle de l'audio de cette vidéo. L'intérêt est de rendre plus facile la recherche d'informations à travers les archives vidéos des réunions passées.

## Chapitre 2

# Personnes et organismes impliqués dans le projet

### 2.1 Maître d'ouvrage

Le maître d'ouvrage sera le contact principal de l'entreprise. C'est avec lui que les étudiants et les tuteurs enseignants seront en relation. Il devra constamment être informé de l'avancement du projet ainsi que des difficultés éventuelles rencontrées. Il pourra interférer à tout moment dans le projet. Son rôle sera de s'assurer que le projet se déroule comme il le désire. Il pourra au besoin demander des modifications ou apporter des précisions sur des points non spécifiés. Toutes les modifications apportées au projet en cours devront être discutées avec les élèves et tuteurs enseignants, pour s'assurer que les objectifs restent adaptés aux contraintes pédagogiques des projets transversaux.

### 2.2 Tuteurs enseignants

Le rôle des tuteurs enseignant est d'encadrer les étudiants tout au long du projet. Ils devront être tenus informés chaque semaine de l'avancement du projet, ainsi que des difficultés rencontrées. Les tuteurs auront pour rôle de guider les étudiants dans la démarche organisationnelle du projet, et de s'assurer que les objectifs fixés sont respectés à la fin de chaque phase de travail.

### 2.3 Elèves Étudiants

Les deux étudiants seront les concepteurs du projet. C'est à eux de mener de bout en bout le projet. Avec l'aide des tuteurs enseignants et du maître d'ouvrage, ils devront mener à bien le projet, de sa phase de recherche à sa phase de conception. Ils devront rendre compte de leur travail chaque semaine, et res-

pecter les différentes contraintes imposées par le cahier des charges (contraintes temporelles, technologiques...).

## **2.4 Utilisateurs du produit**

Les utilisateurs seront les usagers de la plateforme Project2Cloud. Le projet sera ainsi destiné aux personnes ayant créées un compte sur la plateforme, et désirant rechercher une vidéo. En effet, lors de la recherche vidéo, ce sont les fichiers de transcription associés aux vidéos qui seront utilisés.

## Chapitre 3

# Contraintes sur le projet

### 3.1 Contraintes négociables

#### 3.1.1 Types de modules externes utilisés

L'utilisation de modules et logiciels LIBRES est recommandé. Cette mesure est fonction de la quantité et de la qualité des modules accessibles. Dans le cas où les solutions libres ne correspondraient pas aux attentes du client, l'utilisation de modules propriétaires sera envisagée. Une discussion avec les tuteurs et le maître d'ouvrage devra alors être tenue pour trouver le meilleur compromis technique et financier.

#### 3.1.2 Environnement de développement du projet

##### Système d'exploitation

Le projet sera développé pour être utilisé sur une plateforme Linux.

##### Logiciel et langage de programmation

Le langage informatique utilisé est le langage de programmation orienté objet Java, puisque la plateforme déjà existante est elle même codée avec ce langage.

#### 3.1.3 Langage reconnu par les modules de transcription

Le projet développé sera principalement axé sur la langue française. Le module de transcription devrait donc reconnaître des conversations françaises. Dans le cas où les modules français ne seraient pas suffisamment performant ou indisponibles, une solution de traitement en Anglais pourrait être envisagée. Il conviendrait dans ce cas de discuter avec le maître d'ouvrage et les tuteurs enseignants de la suite à donner au projet. Les recherches d'un module français pourront être abandonnées ou se poursuivre tout au long du projet, dans le cas où un nouveau module serait disponible.

### **3.1.4 Fonctionnalités supplémentaires**

Des fonctionnalités supplémentaires pourront être ajoutées en cours du projet. Ces ajouts éventuels seront fonction de l'avancement du projet, de la validation des fonctionnalités prioritaires. Le choix des fonctionnalités supplémentaires sera discuté avec le maître d'ouvrage et les tuteurs enseignants, pour s'assurer que ces nouveautés restent en accord avec la portée pédagogique du projet transversal.

## **3.2 Contraintes non négociables**

### **3.2.1 Contraintes de temps**

Le projet débute après la réunion de lancement qui s'est tenue le 30 septembre 2011, en présence du maître d'ouvrage, des tuteurs et des élèves. Il devra être mené à terme avant le 26 mai 2012. Les étudiants auront à leur disposition une demie-journée par semaine réservée à la mise en œuvre du projet transversal.

### **3.2.2 Environnement de fonctionnement du projet**

La plateforme de communication est actuellement fonctionnelle. Elle est accessible sur internet, via tous les navigateurs web et tous les systèmes d'exploitation (Mac, Windows, Linux).

### **3.2.3 Lieu de fonctionnement**

La plateforme de communication sur laquelle repose le projet est utilisable depuis n'importe quel type d'ordinateur (PC fixe, portable). Par conséquent, et au vu de la mobilité des supports, la plateforme peut être utilisée dans des environnements variés (salle de réunion, lieu public, environnement extérieur...). La contrainte sonore est capitale dans notre projet. Un son ambiant relativement restreint, voir faible est essentiel pour une utilisation optimale du projet.

### **3.2.4 Contraintes matérielles**

L'utilisation du produit se fera par l'intermédiaire d'un micro. Il pourra s'agir de différents types de micro (micro intégré à l'ordinateur, micro externe). De la qualité du micro dépendra la qualité de la transcription. En cas de besoin matériel spécifique au projet, l'entreprise devra être en mesure de le fournir.

### **3.2.5 Quel est le budget affecté au projet ?**

Aucun budget spécifique n'est prévu pour ce projet. Il devra être réalisé dans la mesure du possible avec des modules et des logiciels gratuits, ainsi qu'avec le matériel mis à disposition par l'école.

### **3.3 Faits et hypothèses utiles**

#### **Facteurs influençant le produit, mais qui ne sont pas des contraintes imposées sur les exigences**

Le facteur le plus influent du projet est la qualité sonore. Elle comprend le niveau sonore ambiant autour de l'utilisateur, la qualité du micro utilisé, la prononciation du locuteur. Tous ces facteurs sont indépendants de la conception, et ne constituent pas d'exigences propres au sujet.

#### **Hypothèses que l'équipe fait sur le projet**

L'équipe considère le projet réalisable. La réussite du projet sera fonction de la bonne communication entre les différents protagonistes.



## Chapitre 4

# Exigences fonctionnelles

### 4.1 Portée du travail

#### 4.1.1 La situation actuelle

Project2Cloud est une plateforme actuellement en ligne. Elle est accessible à l'adresse : <http://project2cloud.org/>. Elle est ouverte aux professionnels et aux particuliers.



FIGURE 4.1 – Site de présentation de l'application.

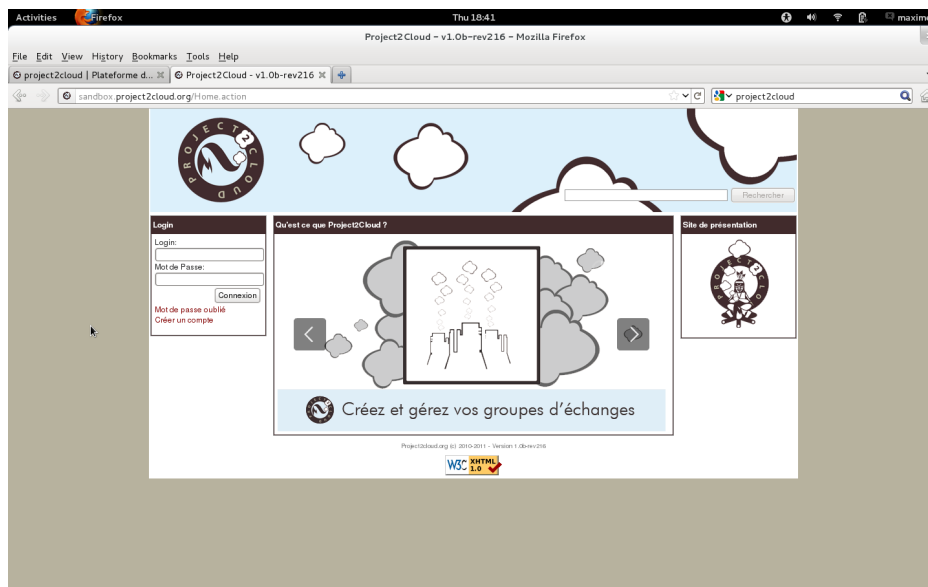


FIGURE 4.2 – Plateforme Project2Cloud.

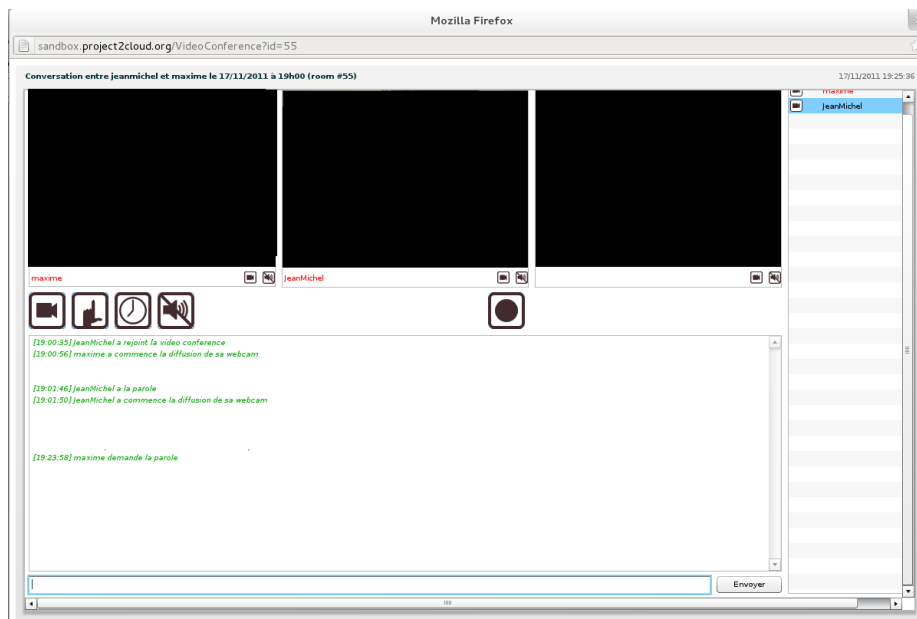


FIGURE 4.3 – Conférence vidéo sous Project2Cloud.

### 4.1.2 Contexte du travail

Notre projet a pour but d'amener une nouvelle fonctionnalité parmi celles citées précédemment. Celle de retranscrire les conversations des sessions, et ainsi de faciliter l'indexation des conférences vidéos.

### 4.1.3 Portée du produit (cas d'utilisations)

Les cas d'utilisation présentés seront ceux en rapport avec notre projet. Il ne prendront pas en compte les cas d'utilisation de la plateforme déjà existante. Pour la partie transcription de texte, notre module devra recevoir le fichier son encodé par la plateforme et le convertir en fichier texte. Il devra également y associer un fichier XML dans lequel les mots clés de la conversation seront indexés grâce une timeline. Une fois le texte transcrit, ce dernier devra être pris en charge par le serveur. Le serveur devra enregistrer le texte ainsi que l'indexation pour en garder la trace.

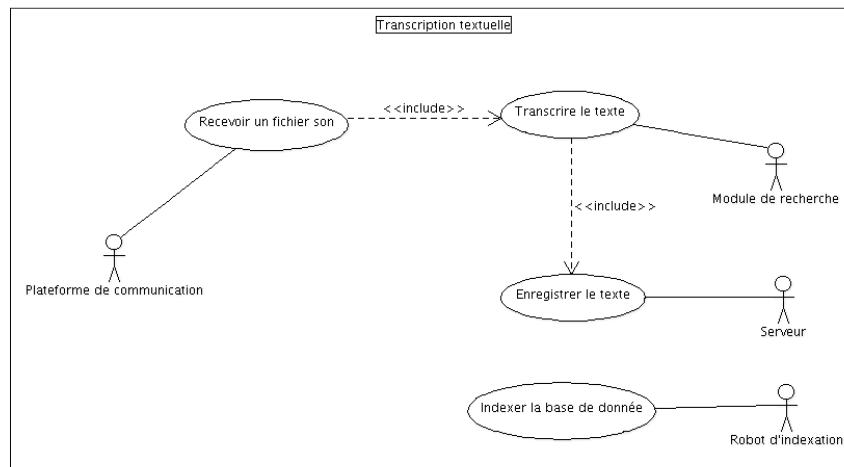


FIGURE 4.4 – Diagramme des cas d'utilisation de la transcription vocale.

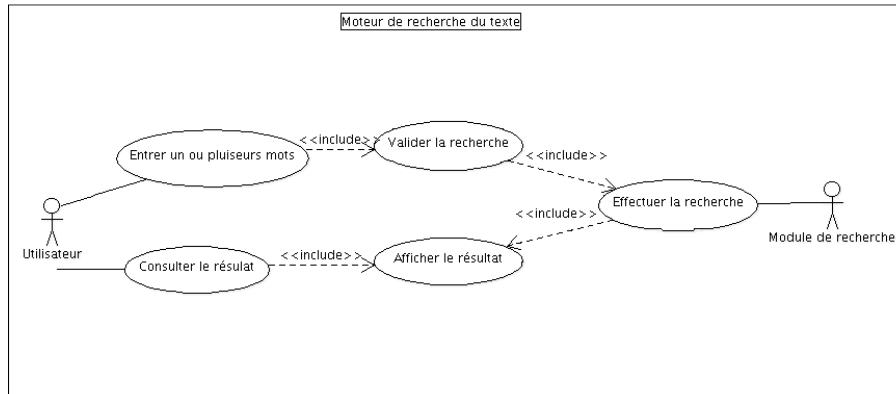


FIGURE 4.5 – Diagramme des cas d'utilisation du moteur de recherche.

## 4.2 Exigences fonctionnelles et exigences sur les données

### 4.2.1 Exigences fonctionnelles

Le projet sera basé sur le traitement de fichiers sonores.

#### Transcription du fichier son en texte

La transcription, ainsi que l'indexation, s'effectueront en tâche de fond. Lors de la réception du fichier son, le module transcrira la voix en texte, et grâce aux modifications apportée, indexera les mots clés dans un second fichier en leur associant leur mappage temporel.

#### Amélioration du modèle de reconnaissance

La Plateforme Project2Cloud donnera la possibilité à l'utilisateur d'améliorer manuellement la transcription d'une vidéo. Lorsqu'une telle correction est disponible, le module de recherche peut alors améliorer son modèle de reconnaissance. Lorsque le modèle aura évolué, il sera intéressant de transcrire de nouveau les anciennes vidéos, pour obtenir une meilleure transcription.

### 4.2.2 Exigences sur les données

Toutes les données mises en jeu dans les modules devront s'adapter avec la plateforme. Elles devront respecter les normes imposées et ne présenter aucun risque de compatibilité. Les formats d'enregistrement et de traitement seront définis plus tard, notamment lors du choix des modules.

## Chapitre 5

# Exigences non fonctionnelles

### 5.0.3 L'interface

L'interface sera simple et répondra au mieux aux contraintes imposées. La partie transcription se fait automatiquement, sans intervention de l'utilisateur. Il n'y a donc pas d'interaction avec le client pour cette partie. La partie interrogation, quant à elle, doit être intuitive et facile d'accès pour n'importe quel type d'utilisateur. L'interface n'a pas encore été décidée, elle le sera plus tard, et sera fonction du temps réservé à cette partie du projet.

### 5.0.4 Le style du produit (packaging inclus)

Le produit livré au client n'aura pas de "style" particulier". Les modules seront développés indépendamment de la plateforme, et y seront rattaché en fin de projet. Il n'est donc pas nécessaire d'inclure une notion de style dans notre projet.

## 5.1 Facilité d'utilisation et facteurs humains

### 5.1.1 Facilité d'utilisation

La partie transcription ne sera pas visible pas l'utilisateur. Il interagira avec la plateforme Project2Cloud comme il le ferait actuellement, puisque la transcription se fera automatiquement, sans aucune intervention de sa part. La facilité d'utilisation de cette partie de notre projet est donc fonction de la facilité d'utilisation actuelle.

### 5.1.2 Personnalisation et internationalisation

Le module ajouté à la plateforme ne sera pas personnalisable par l'utilisateur. L'interface et les fonctionnalités mis en place sur la plateforme ne pourront pas être modifiés. La langue de base de la plateforme est le français. Le site

internet est entièrement rédigé en français et ne propose pour le moment pas d'autre langue. Le projet a pour but de ne reconnaître que des voix françaises, sans proposer d'alternative dans d'autres langues. Il n'y a donc pas de projet d'internationalisation pour le moment.

### **5.1.3 Exigences d'accessibilité**

Le projet étant intégré dans une plateforme déjà existante, les exigences d'accessibilité seront les mêmes qu'actuellement.

## **5.2 Fonctionnement du produit**

### **5.2.1 Rapidité d'exécution et temps de latence**

La transcription n'a pas besoin d'avoir un temps d'exécution rapide puisqu'elle se fera automatiquement sur un serveur.

### **5.2.2 Précision et exactitude**

La transcription se devra d'être la plus précise possible. Le degré de précision dépendra cependant de la qualité du module sélectionné. Éventuellement, plusieurs possibilités de transcriptions pourront être sauvegardées. Le moteur d'interrogation devra prendre en compte que les transcriptions ne sont pas toujours correctes, et donc faire une recherche avec des mots phonétiquement proches.

### **5.2.3 Fiabilité et disponibilité**

Le projet devra être entièrement fiable et répondre au mieux à la demande du client. La qualité de la transcription ne sera pas en rapport avec le bon fonctionnement du module. Aucune erreur grave ne devra survenir pendant la transcription, sous peine de perdre complètement l'intérêt du projet. La partie recherche devra elle aussi être fiable et sans aucune erreur fatale.

### **5.2.4 Robustesse ou tolérance à un emploi erroné**

La qualité de la transcription dépendra du module sélectionné mais également de la qualité de parole de l'utilisateur. Une locution claire et articulée est conseillée pour un résultat optimal. Dans le cas d'une locution médiocre ou d'une mauvaise qualité du son d'entrée, les résultats seront erronés et ne peuvent pas être connus à l'avance. Pour la partie recherche, le module prendra en compte toutes les erreurs potentielles que pourrait commettre l'utilisateur (saisie d'un mot inexistant, validation de la recherche sans mot écrit dans la zone de recherche...). Le module interagira dans ce cas avec l'utilisateur pour lui indiquer la nature de son impair.

### **5.2.5 Capacité de stockage et montée en charge**

A ce stade du projet, nous ne connaissons pas encore les capacités limites de stockage et la montée en charge du projet.

### **5.2.6 Adaptation du produit à une augmentation de volume à traiter**

Le projet devra pouvoir gérer une augmentation de volume d'utilisateurs. Quelque soit le nombre d'utilisateurs connectés simultanément, le module devra être en mesure d'offrir la même qualité de service à tous.

### **5.2.7 Longévité**

Le projet devra être opérationnel sur la durée que le client jugera nécessaire. C'est à lui que reviendra le choix de la longévité du produit. Les modules développés devront aussi être adaptables dans le temps, pour une éventuelle amélioration future. La longévité du projet dépendra donc aussi de sa qualité de réutilisation (code source commenté, documentation explicite et complète...).

## **5.3 Adéquation du produit avec son environnement**

### **5.3.1 Environnement physique prévu**

Le projet sera intégré directement dans une plateforme existante. Il devra donc parfaitement s'intégrer avec le site internet hôte.

### **5.3.2 Environnement technologique prévu**

L'environnement technologique sera uniquement le web.

### **5.3.3 Applications « partenaires » (avec lesquelles le produit doit collaborer)**

Le projet n'a pas pour but de collaborer avec des applications autres que la plateforme hôte qui le reçoit.

#### **5.3.4 Approche « produit » prêt à être commercialisé**

### **5.4 Maintenance, support, portabilité, installation du produit**

#### **5.4.1 Maintenance du produit**

Une fois le produit livré et validé par le client, la maintenance ne sera plus assurée dans le cadre du projet transversal.

#### **5.4.2 Conditions spéciales concernant la maintenance du produit**

Aucune condition spéciale concernant la maintenance n'est à définir dans le cadre du projet.

#### **5.4.3 Installation du système**

A ce stade du projet, nous ne sommes pas en mesure de répondre à cette section du cahier des charges.

### **5.5 Sécurité**

#### **5.5.1 Accès au système**

A ce stade du projet, nous ne sommes pas en mesure de répondre à cette section du cahier des charges.

#### **5.5.2 Intégrité**

A ce stade du projet, nous ne sommes pas en mesure de répondre à cette section du cahier des charges.

#### **5.5.3 Protection des données à caractère personnel**

Les données à caractère personnel ne sont pas directement liées à notre projet. Ces données sont gérées par la plateforme en place.

#### **5.5.4 Audit et traçabilité**

A ce stade du projet, nous ne sommes pas en mesure de répondre à cette section du cahier des charges.



#### **5.5.5 Protection contre les infections**

A ce stade du projet, nous ne sommes pas en mesure de répondre à cette section du cahier des charges.

### **5.6 Exigences culturelles et politiques**

#### **5.6.1 Exigences culturelles**

Aucune exigence culturelle n'est spécifiée pour ce projet.

#### **5.6.2 Exigence politiques**

Aucune exigence politique n'est spécifiée pour ce projet.

### **5.7 Lois et standards influençant le produit**

#### **5.7.1 Conformité avec la loi**

Le projet sera conforme avec la loi et les exigences juridiques française.

#### **5.7.2 Conformité avec des standards**

A ce stade du projet, nous ne sommes pas en mesure de répondre à cette section du cahier des charges.

## Chapitre 6

# Autres aspects du projet

### 6.1 Questions sans réponse

### 6.2 « COTS » : Progiciels ou composants commerciaux

A ce stade du projet, nous ne sommes pas en mesure de répondre à cette section du cahier des charges.

### 6.3 Nouveaux problèmes, créés par l'apparition du nouveau système

A ce stade du projet, nous ne sommes pas en mesure de répondre à cette section du cahier des charges.

### 6.4 Tâches à faire pour livrer le système

A ce stade du projet, nous ne sommes pas en mesure de répondre à cette section du cahier des charges.

### 6.5 Contrôle final de qualité sur site (Cutover)

A ce stade du projet, nous ne sommes pas en mesure de répondre à cette section du cahier des charges.

## **6.6 Risques liés au projet**

Les risques éventuels liés au projet seront découverts au moment de la partie de conception, et dépendront du langage informatique utilisé.

## **6.7 Estimation des coûts du projet**

Le projet n'aura aucun coût spécifique.

## **6.8 Manuel utilisateur et formations**

Les manuels utilisateurs et formations seront conçus au fur et à mesure de l'avancement du projet.

## **6.9 Salle d'attente : idées pour les futures versions**

Les idées pour des versions futures n'ont pas été abordées. Elles le seront dans le cas où l'avancement du projet le permettrait, et résulteraient d'un accord entre les tuteurs enseignants et le maître d'ouvrage.

Troisième partie

Étude bibliographique

# Table des matières

---

<b>1</b>	<b>Transcription automatique d'un fichier audio vers un fichier texte</b>	<b>29</b>
1.1	Principe de fonctionnement de la reconnaissance automatique de la parole . . . . .	29
1.2	Difficultés de l'analyse de la parole . . . . .	29
1.2.1	Décompositions de la paroles en phonèmes . . . . .	30
1.2.2	Variabilité de la parole . . . . .	31
1.3	Variantes du problème de la reconnaissance de la parole . . . . .	32
1.3.1	Locuteurs connu et inconnu . . . . .	32
1.3.2	Vocabulaire connu et limité ou vocabulaire large . . . . .	34
1.3.3	Découpage temporel en locuteurs . . . . .	34
1.3.4	Qualité acoustique de la composante parole . . . . .	35
<b>2</b>	<b>Recherche d'information dans des documents parlés</b>	<b>36</b>
2.1	Son intérêt . . . . .	36
2.2	Sa difficulté . . . . .	36
2.3	Différentes méthodes . . . . .	37

---

# Chapitre 1

## Transcription automatique d'un fichier audio vers un fichier texte

### 1.1 Principe de fonctionnement de la reconnaissance automatique de la parole

La reconnaissance de la parole (RAP) est une forme de reconnaissance de forme, et suit donc le même type de procédure [1]. Ce dernier commence par la capture d'un signal correspondant au signal entrant enregistré par un locuteur lorsque celui-ci parle dans un microphone. Ces données sont par la suite compressées. Cette étape est primordiale pour pouvoir, d'une part travailler dans des temps raisonnables et d'autre part, permettre au processus de reconnaissance de se concentrer sur les aspects pertinent du signal. En effet, il est inutile de garder toutes les données, puisque beaucoup d'entre elles sont inutiles pour la RAP. Cependant, c'est aussi une étape très sensible, car une mauvaise compression peut également amener à perdre des informations primordiales au bon fonctionnement de la RAP. Vient ensuite la recherche de caractéristique dans le signal. On cherche ainsi à trouver un modèle ce rapprochant au maximum de ces caractéristiques. Enfin, il faut prendre une décision sur le modèle qui minimise au mieux et le coût, et les erreurs.

### 1.2 Difficultés de l'analyse de la parole

De nombreux obstacles s'opposent à une bonne reconnaissance automatique de la parole.

### 1.2.1 Décompositions de la paroles en phonèmes

#### Qu'est ce qu'un phonème ?

La parole est une succession de sons, que l'on identifie comme étant des phonèmes. Un phonème est en fait la plus petite unité d'un mot : un mot est constitué de syllabes, qui sont elles même constituées de phonèmes. Par exemple le mot "être" est constitué d'une seule syllabe, en revanche, il est constitué de trois voir quatre phonème (dépendant du fait que le locuteur prononce le 'e' final ou pas). Ces phonèmes sont représentés entre slash, de la même manière que les symboles de phonétique. Dans notre exemple nous avons donc les phonèmes suivant : /ɛ/, /t/, /r/ et éventuellement /e/ .

Le contenu basique de la parole est donc une séquence de phonèmes, qui une fois décodés, sont traduit par une séquence de mot grâce à une correspondance dans un dictionnaire.

#### Caractéristiques des phonèmes

Les phonèmes sont caractérisé par trois paramètres :

- le mode d'articulation
- la place de l'articulation
- la voix

Cependant par la suite, nous ne nous intéresserons dans le cadre de la RAP qu'au mode d'articulation.

Le mode d'articulation correspond au type de son, que nous pouvons classer en 5 groupes majeurs :

- Consonne occlusive : elle fait intervenir un blocage complet de l'air, suivit d'un relâchement soudain (i.e. /p,t,k,b,d,g/)
- Consonne fricative : l'air passe par une fente, formée par le rapprochement des lèvres, des dents, et/ou de la langue avec les lèvres/dents (i.e. /f,s,v,z/)
- Consonne nasale : il y a blocage complet, mais à l'inverse d'une consonne occlusive, l'air passe par le nez (i.e. /m,n/)
- Voyelles : pas d'obstruction de l'air, seule l'emplacement et la forme de la langue et des lèvres modifient le son
- Consonne latérale (ou liquide) et semi-voyelle : le premier correspond aux consonnes pour lesquelles l'air s'échappe par les côtés de la langue, et le second, aux lettres qui donne un son voyelle, mais sont pourtant caractérisées comme étant des consonnes(i.e. /l,r,w,j/)

Il est ici important de rappeler que nous nous intéressons au type de son formé, et non à une lettre. Par exemple, ici la lettre 'c' n'est pas référencée, puisqu'elle se rapporte à deux sons : le son /k/ (e.g. dans le mot "caractère") et le son /s/ (e.g. dans le mot "ceci"). De la même manière, pour le groupe des sons voyelles, il ne faut pas se restreindre aux six voyelles de la langue française, puisque bien d'autres sons de type voyelle existent (e.g. /œ/ dans "oeuf", /ɛ/ dans "lait").

Les systèmes de RAP n'ont pas de grandes difficultés à classer ces groupes de phonèmes, tout du moins lorsqu'il s'agit des groupes de consonnes occlusives,

fricatives et de voyelles. En effet, ces sons ont de fortes caractéristiques, notamment une forte énergie périodique. Or ceci est une information primordiale car la détection de périodes est un facteur majeur de la RAP.

### **La coarticulation**

Un phonème isolé est assez facile à identifier pour un système de RAP. Seulement, lorsque nous parlons, nous ne saccadons pas nos paroles en son. Nous enchainons les phonèmes, les mots, les uns après les autres, d'une manière plus ou moins fluide. C'est dans ce contexte qu'apparaît le principe de coarticulation.

Un phonème intervient de façon significative sur le phonème qui le suit, mais également sur le phonème qui le précède. En effet, l'appareil vocal humain est fait de telle manière qu'il est impossible de passer d'un phonème à un autre (i.e. d'un son à un autre) instantanément. Pour chaque enchainement de phonèmes, il y a un moment de transition (plus ou moins importante et plus ou moins détectable). Pendant celle-ci, la forme des lèvres et la position de la langue et de la mâchoire vont se modifier, pour correspondre au prochain phonème. Par exemple, pour passer d'un son /s/ à un son /t/, la transition est le moment où l'on commence à avancer la langue jusqu'à ce que celle-ci touche les dents. Bien que difficilement détectable pour l'oreille humaine, cette transition apparaît très clairement en terme de signal.

De plus, dans la langue française, les règles de grammaires veulent que nous fassions les liaisons entre les mots, faisant ainsi apparaître des phonèmes.

Ces deux phénomènes de coarticulation et de liaisons rendent la RAP très complexe, puisque un dictionnaire va associer à un mot les phonèmes qui le composent, et non les transitions. Ces transitions ne sont pas identifiables en terme de phonème et sont différentes pour chaque personne (vitesse de parole, appareil vocal différent, etc...). Il faut donc que le système de RAP apprennent à ne pas prendre en compte ces transitions. La liaison quant à elle est un tout autre problème, car elle fait apparaître des phonèmes tout à fait identifiables, "modifiant" ainsi les mots. C'est donc une nouvelle contrainte qui s'ajoute pour la RAP. Il est en effet parfois difficile de distinguer ces liaisons comme étant réellement des liaisons, e.g. "on en" et "on n'en". Nous pouvons en déduire qu'une reconnaissance simple des phonèmes n'est pas suffisante pour une bonne RAP : il faut y associer toutes les règles de grammaire de la langue.

### **1.2.2 Variabilité de la parole**

Beaucoup de variabilité dans l'aspects de la parole peuvent intervenir et ainsi perturber la RAP [2]

- la physiologie de chaque individu : chacun individu est unique, et il en va de même pour son appareil vocal. En conséquence, chaque son produit par un individu est unique. Il faut donc un système de RAP suffisamment performant pour réussir à trouver une ressemblance dans un même phonème prononcé par différente personne.



- Les accents étrangers et régionaux : la langue d’origine influe inévitablement sur la prononciation, ainsi un espagnol par exemple aura tendance à rouler ses ‘r’ en français, ce qui change le phonème, créant une séquence de phonème qui ne peut pas être associé à un mot. De la même manière, pour les accents régionaux, bien que la langue reconnue est leur langue d’origine, ces individus ont un accent qui vient également changer certains phonèmes, et qui vient même en rajouter (e.g. l’accent marseillais). Pour ce cas, il faut un système qui soit capable de repérer que tel ou tel individu a tendance à prononcer ces phonèmes de telle manière.
- La vitesse de parole : tout le monde parle à sa vitesse, et cela complique d’avantage une bonne RAP, puisqu’entre deux individus, un qui parle lentement, et le second qui parle vite, la longueur des phonèmes varie énormément.
- L’état émotionnel : nous sommes humains, nous avons donc toute une panoplie d’émotions qui peuvent faire varier notre parole. Une personne en colère a tendance à parler fort et vite, supprimant ainsi certains phonèmes. Une personne sous pressions, stressée, a la gorge serrée, ce qui modifie les phonèmes.
- etc.

On voit donc ici que beaucoup d’éléments peuvent venir perturber la RAP, même pour une unique personne. Suivant les instants, cette personne peut parler plus ou moins vite, se mettre en colère, etc... Si l’on y rajoute une personne avec un accent, qui est elle-même sujette à toutes ces variations, on se rend compte de l’extrême complexité pour établir un bon système de RAP.

## 1.3 Variantes du problème de la reconnaissance de la parole

### 1.3.1 Locuteurs connu et inconnu

Il existe deux types majeurs de systèmes de RAP :

- Locuteur connu
- Locuteur inconnu

#### Locuteur connu

Ce type de système est le plus répandu, car il est le plus facile à réaliser. Il requiert une période d’entraînement (i.e. de réglage) pendant laquelle il lit des phrases à voix haute des phrases que lui indique le logiciel. Ainsi, ce dernier apprend la manière de parler de l’utilisateur. Il possède une base de données de phrases (qui sont choisies pour leur diversité de phonèmes), et sachant ce que l’utilisateur va dire, il peut associer chaque son à sa signification.

## Locuteur inconnu

Ce deuxième type de système, quant à lui, est immédiatement opérationnel. Il ne requiert aucune période d'entraînement, et apprend seul, au fur et à mesure que l'utilisateur parle, à identifier ses phonèmes.

## Analyse

Lorsque l'utilisateur est connu, le système donne souvent de bien meilleur résultat. Cependant, il présente un inconvénient majeur : une importante perte de temps pour l'utilisateur. De plus, son principe même a de fortes lacunes. En effet, ce type de système fait **lire** l'utilisateur alors que pendant son utilisation, les chances sont grandes que le locuteur parle spontanément plutôt que ne lise. Cependant, il est montré qu'en terme de signal, les phrases lues et les phrases parlées spontanément sont complètement différentes [3]. La spontanéité demande plus de réflexion, et il y a plus d'émotion lorsque nous parlons que lorsque nous lisons. De plus, lorsque nous parlons spontanément, l'espace spectral se réduit, et plus la parole est spontanée, plus la réduction est importante (comme le montre la figure 1.1). Or une réduction de l'espace spectral provoque une réduction de la précision de reconnaissance de phonème, ce qui implique une réduction de la RAP.

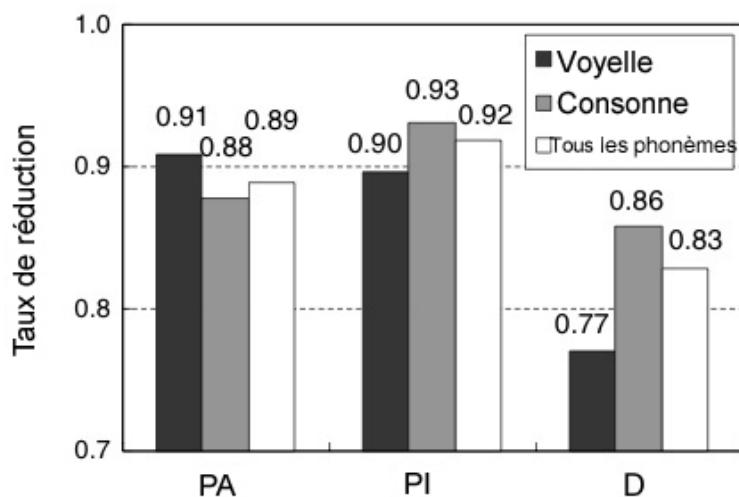


FIGURE 1.1 – Moyenne du taux de réduction des voyelles, consonnes et tous les phonèmes pour les types PA, PI et D de paroles, respectivement Présentation Académique, Présentation Improvisée et Discours, par rapport à des phrases lues. [3]

### 1.3.2 Vocabulaire connu et limité ou vocabulaire large

Pour certains systèmes, le besoin de vocabulaire est très limité, et le système en perd en complexité. Par exemple, les répondeurs à reconnaissance automatique de la parole ne demandent que très peu de mots : "oui", "non", "suivant", "supprimer"... Il sont donc beaucoup plus facile à coder, puisque la suite de phonème ne correspond qu'à un mot, et les possibilités sont très rapides à parcourir.

En l'occurrence, notre projet requiert un vocabulaire très large, puisqu'il doit pouvoir identifier tous les mots qui peuvent survenir lors d'une conversation entre différentes personnes. Cependant, même avec un vocabulaire extrêmement complet, il arrive couramment que des mots inconnus soient prononcés (e.g. nom propres, jargon scientifique, etc.) [4], et si un mot n'est pas dans le dictionnaire, alors la transcription sera forcément fausse. Pour palier à ce problème, plusieurs solutions existent. La première consiste à proposer à l'utilisateur plusieurs hypothèses, avec la possibilité d'ajouter un mot (i.e. de la même façon qu'il est possible d'ajouter des mots aux correcteurs orthographiques de logiciels de bureautique). Néanmoins, cette solution n'est pas envisageable pour notre projet, car ces vidéos seront faites lors de réunions professionnelles, et il n'est donc pas possible d'interrompre l'utilisateur pour chaque mot que le logiciel n'arrive pas à identifier. La deuxième solution, quant à elle, propose de rechercher les mots inconnus dans les documents associés à la vidéo. Cette dernière semble tout à fait appropriée à notre projet, puisqu'il est prévu que pour chaque réunion (i.e. visio-conférence) lui soit associé en plus de la vidéo et de l'audio, les documents utilisés lors de cette dernière. Il serait donc possible de parcourir ces documents afin d'éventuellement trouver le mot correspondant. Cependant, cette proposition ne peut pas fonctionner systématiquement, puisque d'une part, il n'est pas obligatoire que des documents aient été utilisés lors de la réunion, et d'autre part, quand bien même des documents auraient été partagés, les mots inconnus ne sont pas forcément contenu dans ces documents. Pour ce problème, nous pouvons proposer une troisième solution, qui consiste à faire une première passe du contenu parlé avec un tout autre système pour essayer d'en retirer le sujet de la vidéo, et ainsi créer un vocabulaire spécifique à ce sujet. Cette solution pourrait éventuellement faire l'objet d'une amélioration future par l'entreprise, mais elle ne sera pas employée pour notre projet.

### 1.3.3 Découpage temporel en locuteurs

Lorsque plusieurs locuteurs sont présents, il est intéressant de découper le flux de paroles en locuteurs. En effet, un bon modèle s'adapte au locuteur, or si nous ne différencions pas les différents locuteurs, le style de parole changeant à chaque prise de parole, il sera impossible au système de s'adapter convenablement. Ainsi, en découplant en locuteur, le système peut avoir plusieurs "adaptation" pour correspondre à chacun des locuteurs.

Il n'est pas prévu de mettre cette fonctionnalité dans le projet pour l'instant, puisque nous partons du point de vue qu'il n'y a qu'un seul individu de chaque

côté des micros. Dans ce contexte, la division est très simple, puisqu'une entrée correspond à un locuteur. Cependant, à terme, il doit être possible d'avoir plus d'une personne pour un micro, il y a donc une chance que l'entreprise nous demande d'inclure cette fonctionnalité.

#### **1.3.4 Qualité acoustique de la composante parole**

La précision de la RAP peut très vite chuter si la qualité audio est mauvaise, notamment s'il y a des bruits de fond [5], de la musique, une mauvaise transmission, etc.

## Chapitre 2

# Recherche d'information dans des documents parlés

### 2.1 Son intérêt

L'augmentation constante de l'activité informatique et la réduction des coûts de stockage conduisent à un très fort nombre de données de tout types, échangées et stockées. En conséquence, la recherche de données est devenue un domaine clé, tout particulièrement celui de la recherche de données textuelles, qui est le plus courant [4].

La recherche de document parlé, quant à elle, ne s'est pas beaucoup développée au fil des années, notamment dû à son manque de demande et d'exactitude. Cependant, dans les dernières années, l'explosion de contenu audio/vidéo disponible sur internet en fait un important outil. Néanmoins, ce domaine a encore besoin de progrès, car les résultats ne sont pas à la hauteur des exigences des utilisateurs. En effet, l'idéal serait de faire la recherche dans une transcription alignée avec l'audio, ce qui simplifierait le problème en une recherche textuelle. Malheureusement, dans cette hypothèse, le texte devrait être une transcription exacte de l'audio, ce qui nécessiterait une transcription manuelle. Or cette dernière prend du temps, et donc coûte cher. C'est dans ce contexte qu'intervient la recherche d'information des documents parlés.

### 2.2 Sa difficulté

Lorsque l'on crée un système de RAP pour une recherche d'information, le choix du dictionnaire est primordial. En effet, en général, le vocabulaire d'un ASR est assez restreint. Or, si un mot est prononcé et n'est pas dans le dictionnaire, alors le système de RAP ne le reconnaîtra jamais correctement, et les requêtes d'utilisateur contenant ce(s) mot(s) échoueront systématiquement. Malheureusement, les mots ou groupes de mots recherchés sont souvent les moins

courants et les plus spécifiques à un sujet, qui ne sont généralement pas présent dans les dictionnaires des systèmes de RAP.

On en déduit l'importance d'un système de recherche flexible, qui accepte des solutions avec des mots plus ou moins proches.

## 2.3 Différentes méthodes

La méthode la plus classique est évidemment une recherche textuelle classique dans les transcriptions des vidéos. Cependant, une autre méthode suggère d'attaquer le problème sous un autre angle, en ignorant totalement le texte, et de ne s'intéresser qu'à la prononciation. Le système de RAP, à la place de produire un texte de mots, produirait un texte de prononciation phonétique. La requête de l'utilisateur serait ensuite traduite en langage phonétique, et recherchée dans les transcriptions. Le très grand avantage de cette méthode est qu'elle résout le problème de vocabulaire inconnu. De plus, le système de RAP serait nettement plus performant puisqu'il n'aurait qu'à traduire ce qu'il "entend" en phonétique, et non plus à rechercher de correspondance dans son dictionnaire. En revanche, la recherche de mots clés, quant à elle, serait beaucoup plus longue. Or dans notre cas actuelle, la performance du système de RAP n'a pas de réelle importance puisqu'elle se fait en tâche de fond, contrairement à la recherche qui se fait en temps réel.

Une seconde solution utilise les deux méthodes précédentes, combinant à la fois la recherche par mot et la recherche phonétique. L'idée serait de faire une première recherche large (i.e très flexible quant à l'exactitude de la correspondance des mots) en utilisant les mots de l'utilisateur. Viendrait ensuite une seconde passe sur ces résultats, en utilisant la phonétique. Cette méthode impliquerait donc que le système de RAP produise deux transcriptions : une textuelle et une phonétique. Cela ne pose pas de réelle contrainte, puisque comme dit précédemment, la rapidité du système RAP n'a pas d'importance. La recherche, elle, serait plus rapide qu'avec la méthode n'utilisant que la phonétique, puisqu'ici elle ne se fait que sur un nombre de documents préalablement restreint.

Quatrième partie

Procédure de recette

## 1.1 Introduction

Notre projet s'intégrera dans une plateforme déjà existante. Pour le tester, nous devons vérifier qu'il répond aux attentes fonctionnelles, mais qu'il s'intègre également correctement dans la plateforme hôte. Dans ce cahier de recette nous présenterons une batterie de tests visant à faire valider par le client l'intégralité des modules présentés dans le cahier des charges. Nous présenterons ici les tests sous la forme :

Test effectué : *Résultat(s) attendu(s)*

## 1.2 Tests d'intégration du module dans la plateforme

Démarrage d'une vidéo conférence sur le site "Project2Cloud" : *La vidéo conférence doit démarrer sans message d'erreur ni bug.*

Lire un texte contenant les mots clés : "test", "réunion", "projet", "conférence" : *La vidéo conférence se poursuit sans interruption ni ralentissement.*

Couper la vidéo conférence : *La vidéo conférence se termine sans erreur.*

Ouvrir le module de recherche de texte : *Une interface permettant de saisir du texte doit apparaître*

Saisir le mot "réunion" dans la zone de texte : *La zone de texte doit se remplir du mot saisi*

Valider la recherche en cliquant sur le bouton "Rechercher" : *Une nouvelle fenêtre s'ouvre et affiche l'intégralité des occurrences du mot "réunion" trouvés.*



Cinquième partie

Estimation des charges et  
planifications

Sixième partie

**Modélisation**

# Chapitre 1

## Introduction

La phase de conception est la phase permettant de mettre en place les différentes étapes nécessaires à la réalisation de notre projet. Elle inclue des schémas et des diagrammes structurés représentant les étapes et les fonctionnalités de notre future application. La phase de conception est la dernière étape avant la programmation logicielle, elle doit donc être précise et fiable, dans le but de minimiser au maximum les risques d'erreurs d'implémentation.

La phase de conception s'appuie en grande partie sur ce qui a été mis en place dans le cahier des charges de la phase 1. Les différentes réunions avec les différents protagonistes du projet ont permis de fixer les objectifs prioritaires à concevoir en priorité. Dans notre projet, il s'agira donc de trouver un moyen efficace de transcrire la voix en texte, et de pouvoir exploiter cette transcription. Pour cela, il sera nécessaire d'établir des connexions entre notre module et le serveur de la plate-forme Project2Cloud. Nous nous sommes pour cela initié aux techniques de transcription vocale ( Speech-to-text ) au travers de modules libres existant.

Dans ce compte rendu de phase de conception, nous présenterons le module développé sous différents aspects (diagrammes, partons de conception). Nous estimerons également la charge de travail prévue, via la méthode COCOMO. Cet méthode nous donnera des indications sur le volume de travail à fournir en fonction des choix techniques effectués avec les tuteurs enseignants et entreprise.

## Chapitre 2

# Principe général

Notre module sera développé à part, et ne sera pas directement intégré dans la plate-forme P2C. Il devra cependant échanger des informations contenu sur le serveur de la plate-forme. Les vidéos enregistrées sur le serveur devront être traité par notre module, qui en proposera une transcription de type Speech-to-text. La transcription ainsi effectuée sera renvoyée au serveur. Idéalement, un système d'échange module-serveur pourrait être mis en place, permettant un dialogue continu entre les deux instances. Ceci permettrait de rendre le module auto-didacte et d'améliorer significativement la qualité de la transcription proposée.

IMAGE DU PRINCIPE GENERALE

## Chapitre 3

# Transcription

Le module va périodiquement interroger la base de donnée pour détecter les nouveaux fichiers son encodés disponibles. Lorsque tel est le cas, il lancera ce fichier pour le transcrire grâce à son modèle de reconnaissance. De plus, à chaque mot clé reconnu (mot clé entré pas l'utilisateur), il l'indexera dans un fichier XML. Dans ce dernier, seront associés un mot clés et une donnée temporelle, correspondant au moment où ce mot clé a été prononcé dans la vidéo.

A chaque fichier de transcription sera associé un numéro de version. Ainsi, lorsque le modèle de reconnaissance a suffisamment évolué, une nouvelle transcription sera faite.

Diagramme de classe de la transcription et du dialogue avec le serveur

## Chapitre 4

# Amélioration de modèle

# Chapitre 5

## test

Notre application est composée de deux tests : transcription et mappage temporelle de la transcription

### 5.1 tests unitaires

#### 5.1.1 tests boite noire

Septième partie

Réalisation



# Chapitre 1

## a

Le rapport de réalisation est un rapport court mais qui comporte généralement des annexes longues.

Un premier chapitre décrit les outils utilisés pour la réalisation (IDE, serveur de version, logiciel de test, etc).

Un second chapitre montre le résultat du déroulement du logiciel généralement grâce à des snapshots ce qui doit prouver au lecteur ce qui fonctionne et ce qui ne fonctionne pas.

Un troisième chapitre explicite les spécificités qui ont été mises en oeuvre afin de répondre à la modélisation ou justement les écarts avec cette modélisation. Ce chapitre renvoie sur une annexe avec le code

Un quatrième chapitre explicite les tests prévus et donne un aperçu global de leurs résultats. Ce chapitre renvoie sur le document des tests détaillés (cf. chapitre suivant).

Une conclusion donne le résumé du projet et les perspectives à en tirer.

## Huitième partie

### Tests

# Chapitre 1

## a

Ce dossier de tests doit rendre compte de la vérification du logiciel (test unitaire, test d'intégration, recette interne)

Test unitaire :

Test d'intégration :

Recette interne : Avant de livrer le travail au client, vous devrez effectuer vous-même la procédure de recette "en interne". Pour chaque test décrit, vous noterez le résultat obtenu, ainsi que sa conformité par rapport au résultat attendu. Lorsqu'une différence est constatée, vous la commenterez (erreur inattendue, erreur connue mais pas prioritaire, coût présumé de la mise en conformité...).

Neuvième partie

**Manuel d'utilisation**

# Chapitre 1

a

Dixième partie

Dossier de déploiement

# Chapitre 1

## a

Ce dossier de déploiement de l'application doit préciser les modalités de mise en production du logiciel (installation et configuration)

Onzième partie

PV de recette



# Chapitre 1

a

# Conclusion

Cette conclusion doit faire la synthèse de l'état d'avancement du projet, de votre avis par rapport à cet état d'avancement, et décrire brièvement les prochaines étapes du projet.

# Bibliographie

- [1] Douglas O'Shaughnessy. Invited paper : Automatic speech recognition : History, methods and challenges. *Pattern Recognition*, 2008.
- [2] M. Benzeghiba, R. De Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, C. Ris, R. Rose, V. Tyagi, and C. Wellekens. Automatic speech recognition and speech variability : A review. *Speech Communication*, 2007.
- [3] Masanobu Nakamura, Koji Iwano, and Sadaoki Furui. Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech and Language*, 2007.
- [4] Ciprian Chelba, Timothy J. Hazen, and Murat Saraçlar. Retrieval and browsing of spoken content. *Signal Processing*, 2007.
- [5] Yifan Gong. Speech recognition in noisy environments : A survey. *Speech Communication*, 1994.
- [6] Lawrence R. Rabiner and B. H. luang. An introduction to hidden markov models. *IEEE ASSP Magazine*, 1986.
- [7] Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 1989.

-