Système de synthèse de parole

Sommaire

Chapitre 1 Vue générale du système

1.1 Problématique

Pour bien comprendre le problème de synthèse de parole, il est bon de se rapeller d'abord de l'objectif du projet qui est l'automatisation des deux interfaces :

1. L'interface qui lie la sortie du système d'enregistrement avec l'entrée du système de l'annotation.

2.L'interface qui lie la sortie du système d'annotation avec l'entrée du logiciel manager de données Roots.

-Le système doit-il etre optimisé pour une unique format de texte et/ou de l'audio ou est-il destiné à manipuler plusieurs formats?

Sans doute, les systèmes dépendant d'un seul format sont plus facile à développer. Ainsi, l'utilisation de plusieurs formats nécessite le developpement de plus de programmes au sein du système.

***Hypothèse:**

On suppose que le système peut manipuler uniquement les deux formats mp3 et mp4 pour l'audio, et uniquement un format texte brut pour les documents écrits .

-Les système est-il capable de donner un corpus avec le moindre taux d'erreurs meme lors d'un fonctionnement dans des conditions difficiles ?

En effet, de nombreuses variantes peuvent affecter la bonne perfomance du système :

-Réverberations de la voix pendant l'enregistrement

-Qualité du matériel utilisé (micro, carte son,..etc)

-Bande passante fréquentielle limitée (par exemple : ligne téléphonique)

-Elocution inhabituelle (fatigue, stress, émotions..etc)

***Hypothèse:**

On suppose que la cabine utilisée ne permet pas l'évocation de réverberations et que le matériel utilisé est dans des bonnes conditions. On suppose aussi qu'on ne traitera pas des discussions téléphoniques et qu'il y aura une sélection de participants suivant leurs états en effectuant un test psychotechnique avant de commencer l' enregistrement.

Chapitre 2 Sous-systèmes

2.1.Système d'enregistrement de texte et de la parole

2.1.1 Problématique

Pour traiter la problématique de ce système, il faut se rappeler que l'entrée de ce sous-système est incluse dans celle du système tout entier, alors on sera censé dans cette partie de traiter sa sortie uniquement.

-Les sorties du système d'enregistrement sont-elles censées d'etre stockées sous formes de fichiers (fichiers textes et fichiers audio) ou bien seront-elles stockées dans une base de données ?

En fait, si les données sont dans des bases de données, ceci compliquera la tache pour l'experimentateur dans les étapes suivantes.

2.1.2 Hypothèse

On suppose que le stockage se fait en des fichiers que l'experimentateur importera par la suite dans les outils d' annotation.

2.2 Système d'annotation

2.2.1 Problématique

Pour bien apréhender le problème d'annotation, il est mieux d'en comprendre les différents niveaux de complexité.

-Le système reconnait-il les fichiers audio contenants des mots isolés ou des paroles continues?

Evidemment, il est plus simple de reconnaitre des mots isolés bien séparer par des périodes de silence qu'une séquence de mots constituant une phrase. Dans le cas d'une phrase (un paragraphe ou un texte), non seulement les frontières entre les mots seront difficilement détectables, mais aussi la prononciation de chaque mot sera influencé par le mot qui le précede, c'est-à-dire on tombe dans le problème de l'articulation (l'exemple de liason en francais).

En outre, la complexité augmente au cas d'un texte spontané: phrases grammaticales incorrectes, hésitations, faux départs..etc.

2.2.2 Hypothèse

Pour pallier à se problème, on met en hypothèse que le système sait manipuler les paroles continues avec un taux d'erreurs minime et qu'on enregistre des textes lus uniquement.

Ou bien, on suppose qu'on travaille aussi avec des textes spontanés à condition que le système d'annotation soit accompagné d'un système qui adapte la prononciation.

2.3.Système de gestion de données hétérogènes(Roots)

2.3.1 Problématique

-Est il possible d'importer des fichiers textes ou audio annotés manuellement dans Roots?

Le mécanisme qu'on a pour l'heure actuelle consiste à ce que des données orales et écrites passent dans un premier temps par le système d'enregistrement et puis par le système d'annotation, pour qu'ils soient traiter finalement par Roots.

Pourtant, pendant le passage d'un système à un autre des données risquent d'etre perdues ce qui demande une intervention manuelle de la part de l'éxperimentateur pour pallier à ce problème en important les données annotées perdues dans Roots avant qu'il fasse la gestion du corpus tout entier.

2.3.2 Hypothèse:

On suppose qu'il y a pas le risque de perdre des données dans les interfaces entre les sous systèmes.

Chapitre 3 Approche poposée

3.1 Un seul et unique système

Notre système de synthèse de parole atctuelle est composé de trois sous-systèmes comme vu précedemment. L'approche que nous proposons afin d'établir un système plus automatisé et donc un système plus efficace, est de rassembler les trois sous-systèmes en un seul réalisant toutes les fonctionnalités attendues, et ceci en respectant les contraintes suivantes :

1. On rassemble dans un premier temps les deux systèmes d'enregistrement et d'annotation dans un seul système qu'on nomme le système A. Le système A aura comme entrée le texte, la parole (le participant) et l'expérimentateur (instructions pour le participant +Paramètres d'annotation). Et il aura comme sortie, les fichiers des données orales et écrites sous différents formats.

2. Maintenant qu'on dispose des deux systèmes A et Roots, on les rassemble dans un notre systéme B qui aura cette fois-ci comme entrée la sortie du système A plus les paramétres de gestion définis par l'éxperimentateur, et il aura comme sortie le corpus.

3.2 Limites

En automatisant le système de départ, on aura vers la fin un seul système englobant toutes les entités. Le système finale a pour entrée l'experimantateur, c'est-à-dire instructions, paramétres d'enregistrement, paramétres d'annotations et paramétres de gestion de donées hétérogènes. Ceci implique que l'expérimentateur face 4 taches complexes à la fois, ce qui est impossible!

Ainsi, en suivant cette approche un seul expérimentateur ne suffira plus. Il y a une nécessité de faire intervenir plusieurs expérimentateurs qui doivent se mettre d'accord sur la définition des paramétres necessaires.

Chapitre 4 Perspectives

Pour le moment l'expérience s'effectue dans de très bonnes conditions qui empéchent la superposition du signal audio(parole) avec d'autres signaux, ce qui rend les résultats atteints innapliquables sur d'autres conditions d'enregistrement. C'est ainsi que ce modèle doit pouvoir etre developpé de plus prochainement afin de pouvoir l'appliqué dans autres conditions notamment celles de la vie quotidienne.