

Projet Selena

Objectif : Réaliser une tête animatronique avec laquelle nous pouvons tenir une conversation.

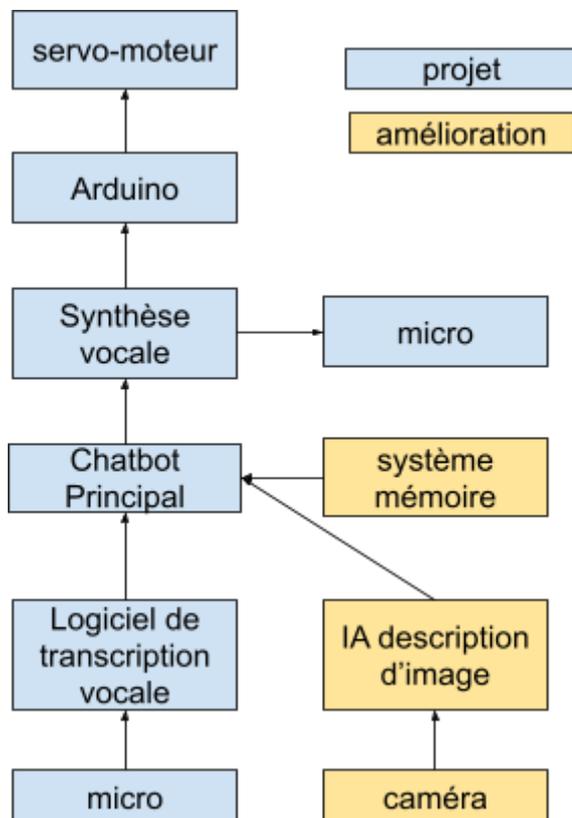
Plus de détails : un microphone sera relié à un logiciel de transcription vocale et transmettra sa sortie au chat bot principal de manière textuel grâce à une balise "dialogue". Le chat bot répondra ensuite toujours de manière textuelle avec d'autres balises telles que "émotion" et "dialogue". La balise dialogue sera renvoyée au logiciel de synthèse vocale. (Probablement espeak, qui n'a pas une bonne qualité mais qui a tout de même la possibilité de générer des fichiers phonèmes) La sortie de la synthèse se fera d'une part sur le système arduino et d'autre part sur le micro.

La balise émotion sera envoyée à l'arduino pour être réalisée sur les servo moteur du robot par l'intermédiaire de paternes. Le fichier phonèmes envoyé à l'arduino permettra de faire de la synchronisation labiale avec le texte émis.

Des améliorations telles que la mémoire ou la vue peuvent être ajoutées grâce au balisé mémoire et vue.

Pour réaliser la vue, nous utilisons une caméra qui prend des photos à intervalles réguliers. Ces photos sont envoyées à une IA de description d'images. Cette description est envoyée dans le prompt du chatbot principal grâce à la balise caméra".

Schéma du fonctionnement:



Brouillon de l'organisation de projet:

Tâches à effectuer:

Software:

1- IA reconnaissance d'image

- a) Trouver le logiciel à utiliser
- b) Apprendre à l'utiliser pour retourner une string
- c) Récupérer le flux vidéo dans le logiciel

2- Mémoire WIKI

- a) Connaître les outils à utiliser.
- b) Réaliser le wiki.
- c) Chatbot qui analyse les dialogues et ressort les informations importantes.
- d) Wiki retourne les strings en lien avec les infos importantes.
- e) Les informations importantes sont triées et écrites dans le wiki.

3- Transcripteur vocal

- a) Connaître les outils à utiliser
- b) Récupérer le flux audio des micros
- c) Retourner une string contenant les phrases dites.

4- Synthétiseur vocal

- a) Connaître les outils à utiliser
- b) Réussir à lui envoyer une string.
- c) Transférer le flux audio vers le haut-parleur
- d) Trouver une voix sympa

5- Arduino expression

- a) Réaliser un panel d'émotion différents avec les valeurs associés pour les servomoteurs
- b) Réussir à envoyer des string à l'arduino
- c) Créer un panel de mouvements réalisables

Les tâches en bleu sont des idées d'améliorations et seront réalisées uniquement si on a le temps et le matériel

Hardware:

1- Imprimer les pièces en 3d

2- Commander le matériel pour la tête

3- Les assembler

4- Réaliser le circuit d'alimentation des servomoteurs et câblage arduino

5- Réaliser un masque

6- Aspect esthétique: lentilles, perruque

- 7- Câble les caméras.
- 8- Câbler les microphones.
- 9- Câbler l'haut-parleur.
- 10- Rajouter des leds, un corps, vêtement...

1- a) Trouver le logiciel à utiliser

J'ai trouvé un modèle de langage qui se nomme mini gpt 4. Il peut générer des réponses par l'analyse d'une image.

b) Apprendre à l'utiliser

Il m'est compliqué de faire tourner le logiciel en raison d'un manque de mémoire de mon gpu. Idéalement, j'ai besoin de 23 Go de mémoire GPU, mais nous pouvons nous limiter à 12. Je vais essayer de faire des tests avec l'api pour voir si les délais sont tenables.

4- Synthétiseur vocal

- a) J'aimerais utiliser Espeak pour décomposer les textes en phonème et ensuite Mbrola pour la synthèse des phonèmes. J'ai trouvé TTS maker, un site où la qualité de la synthèse est impressionnante ! Je peux avoir la clef API gratuite si je travaille en recherche. Il faut que je constitue un dossier pour Selena. Même si la qualité est impressionnante, je vais devoir utiliser speak car tts maker ne me permettra pas de faire des fichiers phonèmes.
- b) Il s'utilise en ligne de commande donc rien de plus simple. Il stocke le résultat dans le fichier passé en paramètre.
- c) On le fera tourner sur raspberry et le haut-parleur brancher sur le port jack.

Idée:

Il va falloir permettre à Selena de poser les questions à son propre module caméra. Pour décrire si elle souhaite quelque chose qu'elle a préalablement vu, ou lire des textes. demande trop de ressources.