



UPSSITECH 3A Systèmes Robotiques et Interactifs

IHM : Conception d'un moteur de fusion multimodale

Rapport de TP

Rédigé par :
Loïc Maignan
Kyllian Aloy

Table des matières

IHM :	1
Conception d'un moteur de fusion multimodale	1
Rapport de TP	1
I- Objectif :	3
II- Architecture :	4
III- Reconnaissance de parole :	5
A) Grammaire	5
B) Sra5	5
IV- Reconnaissance de geste avec ICAR :	6
V- Moteur multimodale	7
A) Introduction	7
B) Chronogrammes	8

I- Objectif :

Le but de ce bureau d'étude est de spécifier, concevoir et implémenter un moteur de fusion multimodale pour interagir avec une palette de dessin ne disposant d'aucun bouton. Pour créer et déplacer des formes sur la palette on utilise les modalités suivantes :

1. La reconnaissance de parole grâce au moteur de reconnaissance de parole (avec l'usage de l'agent ivy sra5)
2. La reconnaissance de geste grâce à la palette de reconnaissance de geste 2D (avec l'usage d'ICAR)
3. Le pointage (souris) sur la palette de dessin

L'objectif est de développer un moteur de fusion des différentes modalités permettant d'approcher le célèbre « put that there », une des premières techniques d'interaction multimodale proposée par le MIT il y a maintenant une quarantaine d'années.

II- Architecture :

Les outils communiquent de la façon suivante avec sra5 comme outil de reconnaissance de parole, ICAR comme outil de reconnaissance de geste et un script processing pour l’affichage graphique des formes. Le moteur multimodal est une machine à état écrit en python.

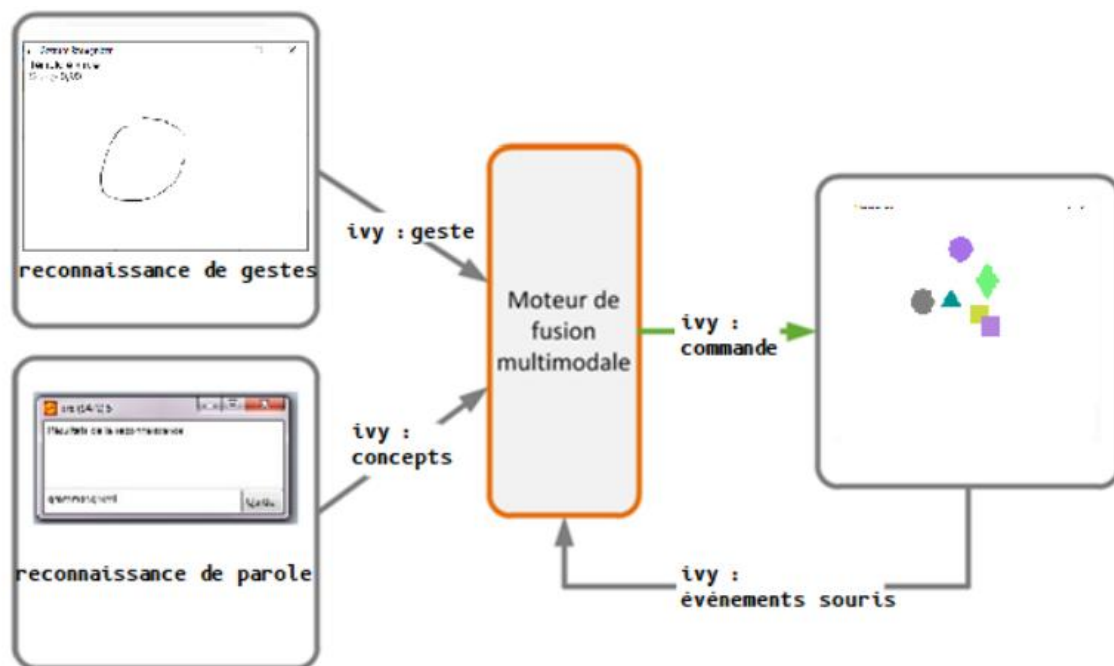


Figure 1: Organisation des outils

III- Reconnaissance de parole :

A) Grammaire

Pour pouvoir effectuer les tâches de création, déplacement et suppression de forme, nous avons créé une grammaire capable de comprendre différents cas d'utilisation.

- Les différentes actions : déplacer, supprimer, créer, ajouter.
- Les différentes couleurs : Rouge, Vert et Bleu.
- Les formes : Triangle, Losange, Rectangle et Cercle
- Les expressions indiquant les lieux : ici, là, ça ...

Grâce à cette grammaire, nous sommes capables de détecter les différentes formes que l'utilisateur peut dire.

- Crée cercle rouge ici
- Supprime ça
- Déplace ça ici

B) Sra5

Par la suite, l'outil sra5 nous permet d'utiliser notre grammaire. Un microphone est d'envoyer les informations reconnues à Ivy, notre centre de messagerie. Pour simplifier le traitement dans notre moteur de recherche, nous avons étiqueté les données qui sont envoyées. Par exemple, si l'utilisateur dit « Crée un cercle rouge ici », nous envoyons : Action = Crée, Forme = Cercle, Couleur = rouge, Lieu = ici

IV- Reconnaissance de geste avec ICAR :

L'outil ICAR est un outil permettant de faire de la reconnaissance de geste dans une fenêtre. Nous pouvons également créer notre propre mémoire de geste pour reconnaître les formes souhaitées.

Dans ce logiciel nous avons donc mémorisé 4 formes que l'utilisateur pourra reproduire dans l'interface utilisateur. Il y a le triangle, le rectangle, le cercle et le losange. De plus, nous l'avons relié à Ivy pour pouvoir faire transiter des messages entre les différents outils.

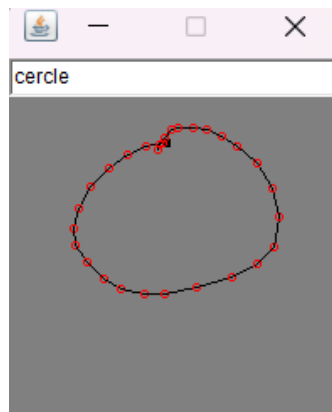


Figure 2: Interface d'ICAR

V- Moteur multimodale

A) Introduction

Pour pouvoir fusionner les informations des différents outils et effectuée des taches en conséquence, nous avons programmé une machine à état en python.

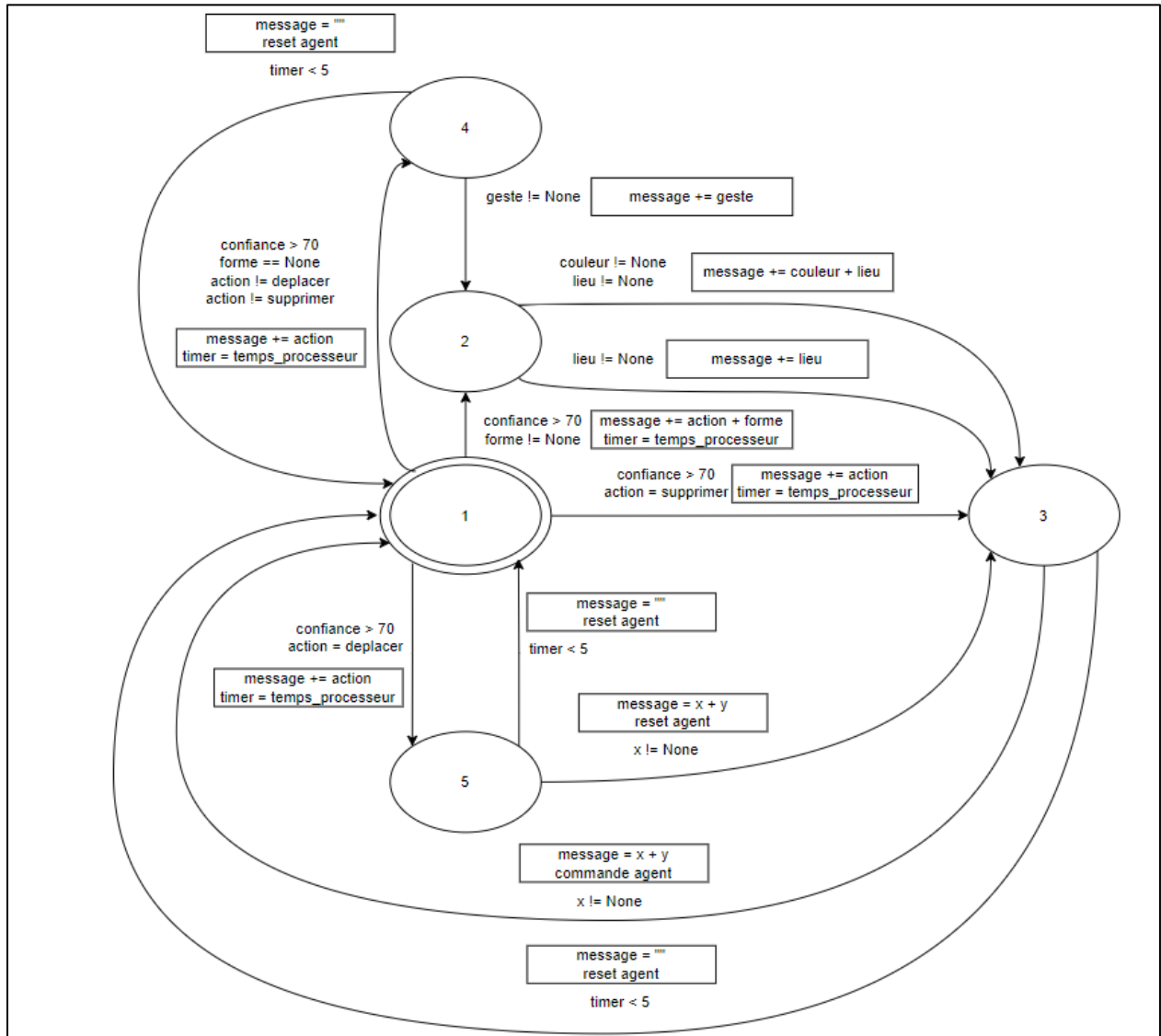


Figure 3 : Machine à Etat du moteur multimodale

B) Chronogrammes

Les chronogrammes suivants illustrent le fonctionnement du système. Voici quelques éléments clés pour les comprendre :

Récupération multimodale des informations

- La collecte des informations (parole, clic sur palette, gestes) s'effectue en parallèle.
- Dans certains cas, la parole peut être dissociée de l'action : elle peut être prononcée avant ou après cette dernière. Les détails spécifiques à chaque situation seront expliqués dans les chronogrammes correspondants.

Utilisation des éléments facultatifs

- Les informations entre parenthèses sont optionnelles : leur présence ou leur absence n'affecte pas le fonctionnement du système, mais elles ajoutent un confort pour l'utilisateur.

Structure des commandes vocales

- Le mot d'action (comme *créer*, *déplacer*, *supprimer*) doit impérativement être prononcé en premier pour que la commande soit reconnue.
- La suite de la phrase, qui contient les détails de l'action, doit obligatoirement être énoncée après le mot d'action. Si ce n'est pas le cas, la commande ne sera pas prise en compte.

1) Action Créer

a.) Premier cas

Cet exemple permet de créer un triangle vert à l'endroit où l'utilisateur clique.

- La phrase peut être prononcée de manière segmentée, c'est-à-dire que les mots peuvent être dits à différents moments, tant que le clic intervient à la fin de la phrase complète. Si le clic est effectué au milieu de la phrase, le système reste en attente de la suite.
- La couleur est facultative : si elle n'est pas précisée, la forme créée sera noire. Cependant, il est impossible de demander explicitement une forme noire.

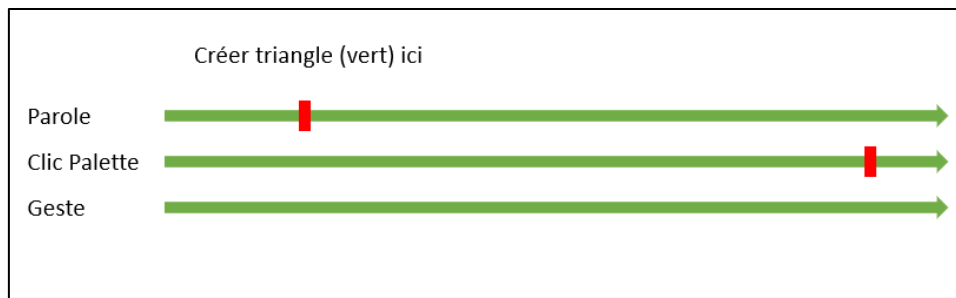


Figure 4 : Chronogramme 1

b) Deuxième cas

Cet exemple crée une forme reconnue et dessinée en rouge à l'endroit du clic.

- Ici encore, la couleur est facultative et sera noire si elle n'est pas spécifiée.
- Contrairement au premier cas, toute la phrase peut être prononcée d'un seul coup sans poser de problème.

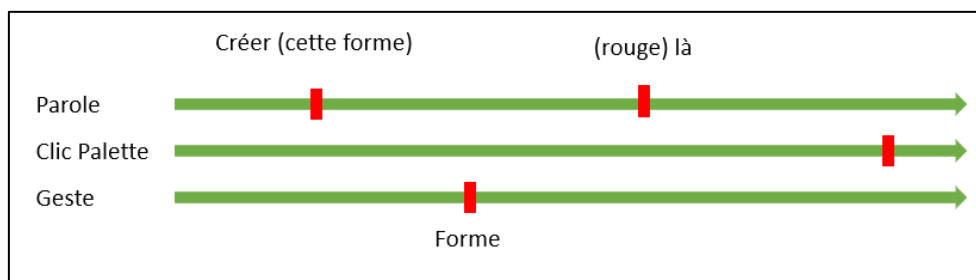


Figure 5 : Chronogramme 2

2) Action Déplacer

Cet exemple modifie la position d'une forme en la déplaçant de son emplacement initial au second endroit cliqué.

- Les expressions entre parenthèses peuvent être énoncées successivement ou segmentées, sans que cela impacte le fonctionnement.
- Tout ce qui est entre parenthèses n'est pas pris en compte par le système et sert uniquement à améliorer le confort de l'utilisateur.

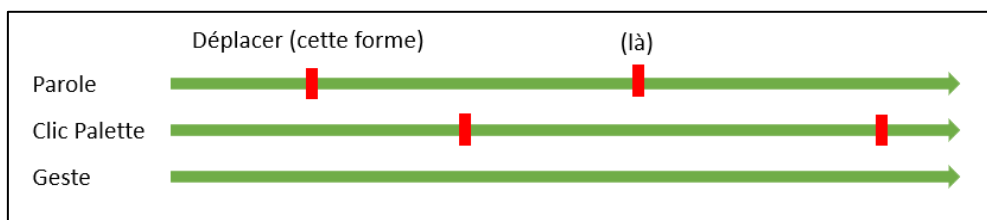


Figure 6 : Chronogramme 3

3) Action Supprimer

Cet exemple supprime la forme cliquée, quelle qu'elle soit.

- Si une forme chevauche une autre, celle au premier plan est supprimée en priorité. Cependant, si vous cliquez précisément sur une partie visible de la forme située derrière, celle-ci sera correctement supprimée.

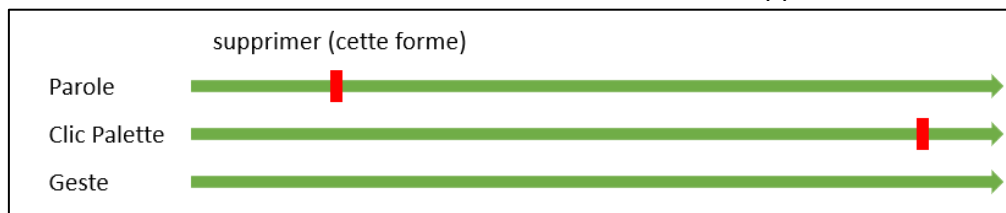


Figure 7 : Chronogramme 4

VI- Résultats

Une fois tous les outils implémenter et connecter, notre application est capable de dessiner, effacer ou déplacer une forme sur l'interface comme le souhaite l'utilisateur.

Amélioration possible :

- Ajout de plus de couleur que rouge vert et bleu (rose, marron, orange ...)
- Ajout d'options de sélection plus large (exemple : Supprime tous les triangles, Efface tous ...)
- Ajout d'autre forme 2D (exemple : pentagone, ellipse ...)
- Ajout de forme 3D (exemple : cube, pavé, sphère ...)