

Travaux Pratiques Interaction Vocale

(Ph. Truillet)
septembre 2019

1. tâche à effectuer

Nous souhaitons concevoir et réaliser une application vocale (**entrée** et **sortie**) permettant à un utilisateur de déplacer à la voix (en entrée – reconnaissance vocale- **ET** en sortie -synthèse vocale-) une forme affichée sur un l'écran (**dans une application que vous devrez coder dans le langage que vous voulez**) dans **au moins** quatre directions : **haut, bas, gauche** et **droite** et **dans les diagonales** 😊.

Il devra être possible de **ramener la forme au centre de l'écran** de manière strictement vocale.

En outre, un retour (**feedback**) **vocal** et/ou **sonore** devra au moins être effectué lorsque la forme atteint un des bords de l'écran.

Afin de réaliser notre application multimédia, **nous nous servirons prioritairement du middleware** (bus logiciel) ivy [<http://svn.tls.cena.fr/wsvn/ivy> et <http://www.tls.cena.fr/products/ivy>], support au TP sur la multimodalité.

Nota : Si vous êtes sous Linux ou MacOS, il vous faudra trouver des solutions alternatives (par exemple, utiliser **MaryTTS** pour la synthèse vocale ou **STT** ou python pour la reconnaissance vocale – cf. liens plus bas).

2. travail attendu de cette séance (2 h)

Après avoir pris en main les agents de reconnaissance et de synthèse vocale avec ivy, l'objet de cette séance est :

1. de **définir la grammaire de reconnaissance** (commandes vocales ou langage « *pseudo-naturel* ») qui sera utilisé par votre application, gérer les résultats sémantiques (i.e. les concepts associés) et le taux de confiance.
2. de **définir les retours vocaux à synthétiser et sonores** utilisés par votre application.
3. de développer une application d'affichage de la forme à l'écran (en java, Processing, python ... ou un autre langage).
4. et enfin **développer le contrôleur de dialogue à l'aide d'une machine à états** (qui peut être inclus dans l'application d'affichage de la forme) basé sur un échange de messages ivy avec au moins les modules de reconnaissance et de synthèse vocale.



A la fin de la séance, vous aurez produit **un prototype haute-fidélité** du système demandé.

Nota : pour ce faire, vous pourrez utiliser quelques agents ivy déjà codés présentés en annexe.

3. téléchargements

- **ppilot5** (Text-to-Speech), **sra5** (Automatic Speech Recognition), ... :
<https://github.com/truillet/upssitech/wiki/3ASRI>
- **librairie java ivy (avec une version de java >= 1.8)** :
<https://github.com/truillet/upssitech/wiki/3ASRI>
- Si vous le désirez, vous pouvez aussi utiliser **MaryTTS** (<http://mary.dfki.de/>), serveur Test-to-Speech écrit en Java
- **STT** : Speech Recognition for Java/Processing basé sur Google Chrome et websockets :
<http://florianschulz.info/stt>
Vous pouvez utiliser la page <https://www.irit.fr/~Philippe.Truillet/stt.html> pour lancer le serveur de reconnaissance.
- **SpeechRecognition**, librairie en Python : <https://pythonprogramminglanguage.com/speech-recognition/>

N'hésitez pas à me demander si tel ou tel agent existe : c'est peut-être le cas ! Et puis, vous pouvez **CODER** vos propres agents selon **VOS** désirs ! 😊

Annexe 1 - utiliser sra5

sra5 est un agent SAPI 5.x utilisant le moteur de reconnaissance natif de Windows Vista, 7, 8.1 ou 10 et peut renvoyer **deux types de solutions** issues de la reconnaissance **sous deux formats différents** :

Lancement de l'agent

```
sra5 -b 127.255.255.255:2010 -p on -g grammaire.grxml
```

- b adresse IP + port
- p mode de renvoi des données (mode parsage **on** ou **off**)
- g fichier de grammaire utilisé (grammaire de type grxml
- cf. <http://www.w3.org/TR/speech-grammar>)

Retours (sur le bus ivy)

- **sra5 Text**=chaîne_orthographique **Confidence**=taux_de_confiance (si le flag *parse* est positionné à off)
- **sra5 Parsed**=resultat **Confidence**=taux_de_confiance **NP=xx Num_A=xx** où NP est le numéro du résultat courant et Num_A le numéro d'alternative (si le flag *parse* est positionné à on)
- **sra5 Event**={Grammar_Loaded | Speech_Rejected} : envoi d'événements provenant du moteur de reconnaissance.

Commandes (sur le bus ivy)

- **sra5 -p {on | off}** sra5 change le mode de retour de la reconnaissance (on → mode de retour sous forme de concept ou off → mode de retour orthographique)
- **sra5 -g** sra5 active une nouvelle grammaire (sur un chemin local à la machine)

Annexe 2 - utiliser ppilot5

ppilot5 permet d'utiliser des systèmes de synthèse vocale compatibles SAPI5.

Lancement de l'agent

```
ppilot5 -b 192.168.0.255:2010 -r Virginie -o "NomDuMoteurTTS"
```

Par défaut, **ppilot5** prend le premier moteur de TTS trouvé et apparaît sur le bus ivy sous le **ppilot5**

- b adresse IP + port
- r nom sous lequel apparaîtra l'agent sous ivy (dans l'exemple précédent, « Virginie »)
- o nom du moteur de synthèse utilisé (difficile à deviner !)

Commandes (sur le bus ivy)

* Synthèse

- **ppilot5 Say=hello** **ppilot5** prononce la chaîne de caractères "hello"

* Commandes

- **ppilot5 Command=Stop** la synthèse vocale est stoppée. **ppilot5** renvoie **ppilot Answer=Stopped**
- **ppilot5 Command=Pause** la synthèse vocale est mise en pause. **ppilot5** renvoie **ppilot Answer=Paused**
- **ppilot5 Command=Resume** la synthèse vocale est relancée si elle était en pause précédemment. **ppilot5** renvoie **w**
- **ppilot5 Command=Quit** l'application se ferme

* Paramètres

- **ppilot5 Param=Pitch:value** le pitch est changé par la valeur donnée. **ppilot5** renvoie **ppilot Answer=PitchValueSet:value**
- **ppilot5 Param=Speed:value** la vitesse est changée par la valeur donnée. **ppilot5** renvoie **ppilot Answer=SpeedValueSet:value**
- **ppilot5 Param=Volume:value** le volume est changé par la valeur donnée. **ppilot5** renvoie **ppilot Answer=VolumeValueSet:value**