

# MCS

# RAPPORT TP 1-3

*Commande de drone par reconnaissance vocale*

**ROUSSEL Hugo**

2018-10-17

M1 info IARF TPA12

## INTRODUCTION

L'objectif est de reconnaître des ordres oraux pour commander un drone. L'algo de Dynamic Time Warping est utilisé pour mesurer la similarité entre deux enregistrements.

## RECONNAISSANCE BASIQUE

Reconnaissance vocale par distance DTW sans détection des silences et sans réduction du bruit.

### PREAMBULE

1. Des enregistrements supplémentaires ont été réalisés (H01, H02, H03, avec H pour Hugo) afin de tester la reconnaissance quand le locuteur référence et le même que le locuteur hypothèse. Les silences de début et de fin ont été supprimés sur certains enregistrements (H01, H02) pour pouvoir tester leurs impacts sur la reconnaissance.
2. Les labels (ordre de drone) sont extraits des noms de fichier, on les suppose correct.
3. La fonction "matriceConfusion" s'occupe des tests, d'afficher les résultats de la reconnaissance, et de produire la matrice de confusion.

### TESTS ET DONNÉES

Données prises du corpus non bruité fournies pour Test 1 et 2. Les enregistrements pour le Test 3 sont fournis dans le dossier "corpus". Le Test 4 est un mélange des deux.

	Hypothèses	Références
Test 1	M01, M02 (24r)	M03 à M013 (143r)
Test 2	M01, M02, F03, F04 (52r)	Corpus complet privé des hypothèses. (182r)
Test 3 (enregistrement personnel)	H02 (silences supprimés), H03 (10r)	H01 (silences supprimés) (5r)

Test 4 (mix)	H02 (silences supprimés), H03 (10r)	M01 à M13 (167r)
--------------	--	------------------

*Liste des tests avec les données utilisées. r = record, nombre d'enregistrement*

Le silence de début/fin (quand présent) est de taille variable. Pour le corpus fournis, ce silence est sans bruit, ce qui n'est pas le cas de mes enregistrements, pour lesquels un petit bruit est produit par le micro.

Le matériel utilisé pour enregistrer (et/ou traiter) les enregistrements du corpus fournis est inconnu, mais semble consistant entre tous les enregistrements.

## RÉSULTATS

Les tests avec leurs résultats sont donnés dans le dossier "Tests" à la racine du projet.

La matrice de confusion produite par les tests est au format csv et peut être importée dans un tableur (les lignes/colonnes de ces matrices ne sont pas toujours dans le bon ordre, la diagonale peut être dure à lire).

	Test 1	Test 2	Test 3	Test 4
Taux d'erreur :	0 %	20 %	40 %	60%

Il y a 13 ordres (labels) différents, on a donc 1 chance sur 13 (environ 8%) de reconnaître l'ordre par hasard, avec l'exception du test 3, seulement 5 ordres ( 20% ). Tous les résultats sont meilleurs que le hasard.

Le Test 1 montre que la méthode donne de très bon résultat quand certains paramètres extérieurs (bruits, tonalité) sont contrôlés. L'ajout d'un facteur de diversité supplémentaire (tonalité) fait baisser les résultats (Test 2).

En regardant le détail du Test 3, on remarque que ce sont les enregistrements dont les silences n'ont pas été supprimés (H03) qui posent problème. En effet, si on ignore H03, le taux d'erreur est de 0%. La suppression automatique des silences pourrait être une bonne approche pour améliorer la reconnaissance dans ce cas.

Pour le Test 4 on aurait pu s'attendre à ce que la principale source d'erreur soit les enregistrements sans silences (H02) mais dans les détails on constate que l'erreur est équitablement répartie (2 reconnus pour H02 et 2 reconnus pour H03). L'erreur pourrait alors être due à la différence du matériel et à la présence de bruit.

## CONCLUSION

La reconnaissance est efficace si les enregistrements testés sont produit dans les mêmes conditions que les enregistrements de référence. L'utilisation de méthode automatique pour réduire le bruit et supprimer les silences devrait permettre d'améliorer significativement les performances dans les cas similaires aux tests 3 et 4.

## NOTES ET REMARQUES

- Le travail sur la suppression des silences à été commencé mais pas continué, faute de temps, l'ajout de tests pour le rapport ayant été privilégié.
- Le code de test initialement fournis produisait une très mauvaise reconnaissance. Après application des correctifs, la reconnaissance est passée de 90% d'échec à 0% ou 20% ! Le problème venait de "MFCCLength" incorrectement utilisé comme taille des tableaux destinés à produire les Field.
- Lancer le "main" présent dans "myDTWtest.java" exécutera les 4 tests à la suite, pour n'en lancer qu'un on peut modifier la liste "tests".