



Travaux Pratiques n°5 & 6

Reconnaissance de la parole par DTW

La Reconnaissance Automatique de la Parole a pour objectif de retrouver le message prononcé dans un signal acoustique. Ce message peut correspondre à une séquence de phonèmes, à un mot, à une phrase, etc.

Dans cette étude vous réaliserez une RAP en mots isolés dans un contexte de vocabulaire restreint.

La méthode de reconnaissance utilisée est fondée sur un alignement temporel dynamique ou programmation dynamique.

1. Système de reconnaissance en mots isolés par DTW

Le système de reconnaissance en mots isolés a pour but de mettre en correspondance un signal acoustique inconnu O avec un dictionnaire de références $\{R_1, \dots, R_n\}$. La reconnaissance est achevée en retenant le mot du dictionnaire le plus proche du signal inconnu au sens du critère de distance D choisie.

L'algorithme d'alignement temporel dynamique (ou programmation dynamique) calcule le chemin permettant d'aligner temporellement deux séquences numériques distinctes ainsi que la distance séparant les deux séquences.

Soit $O = \{O_1, \dots, O_K\}$ et $R = \{r_1, \dots, r_p\}$ deux séquences numériques de longueurs différentes. Dans un premier temps la matrice des distances locales entre les deux vecteurs est calculée pour chaque couple (o_i, r_j) . Dans la pratique la distance euclidienne est souvent utilisée.

A partir de la matrice de distances locales, l'algorithme de programmation dynamique établit le parcours minimisant la distance totale entre les vecteurs O et R selon le critère d'optimisation suivant :

$$c_{ij} = \min \begin{cases} c_{i,j-1} + d_{i,j} \\ c_{i-1,j-1} + 2 * d_{i,j} \\ c_{i-1,j} + d_{i,j} \end{cases}$$

C'est à dire que la minimisation du chemin total peut être décomposée en succession de minimisations locales.

2. Mise en œuvre d'un système d'assistance au contrôle aérien

2.1 Présentation du système

Des efforts sont réalisés depuis les années 90 afin d'intégrer des systèmes de reconnaissance vocale dans les systèmes de contrôle de trafic aérien, notamment afin de reconnaître automatiquement les indicatifs aériens des avions en vol.

Le système de reconnaissance en mots isolés que vous allez mettre en œuvre se décompose en :

- une étape d'extraction des vecteurs de paramètres acoustiques;
- une DTW qui permet de calculer la distance entre le mot prononcé et un mot du dictionnaire;
- une étape de décision ;
- l'affichage du mot reconnu sur le terminal de l'utilisateur;
- la possibilité de visualiser les résultats de la DTW sur l'ensemble du dictionnaire afin d'évaluer la robustesse de la reconnaissance.

2.2 Dictionnaire

Le dictionnaire est composé des lettres de l'alphabet radio international. Nous nous limiterons aux douze premières lettres de l'alphabet : {**alpha, bravo, charlie, delta, echo, foxtrot, golf, hotel, india, juliett, kilo, mambo**}.

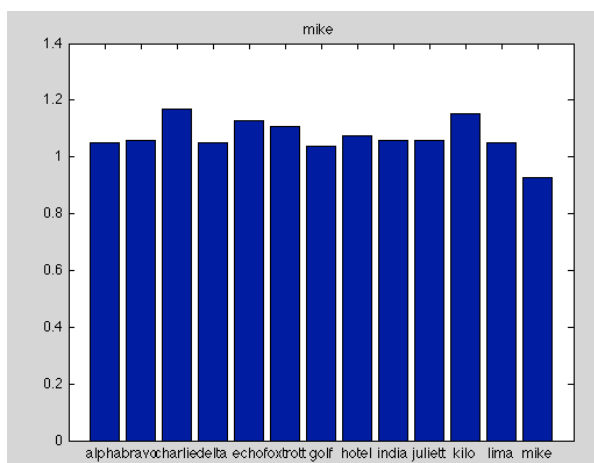
2.3 Spécifications

Vous développerez ce système sous Matlab, en prenant en compte les spécifications suivantes :

- les signaux d'entrée et du dictionnaire vous sont fournis au format wav, ainsi vous travaillerez tous avec les mêmes signaux audio.
- les paramètres acoustiques sont des paramètres cepstraux calculés sur des fenêtres de 30ms toutes les 15ms à l'aide de la fonction `rceps()` de Matlab. Les fenêtres glissantes sont gérées grâce à la fonction `buffer()`. La fonction réalisant l'extraction des vecteurs acoustiques **`extractionCoeffCepstraux()`** vous est fournie.
- les vecteurs acoustiques des signaux de référence ne sont calculés qu'une seule fois puis stockés dans une matrice.

2.4 Réalisation

On vous fournit une trame de programme **`recoDTW_IAN.m`** qui effectue: 1) la paramétrisation de chaque mot (en référence et en test), 2) l'appel de la fonction **`dtw_IAN`** pour chaque mot à reconnaître, 3) la visualisation sous forme d'histogramme de la robustesse de la reconnaissance en comparant les mesures des distances calculées par la DWT (cf. figure ci-dessous pour le résultat du mot testé 'mike' : la plus petite distance obtenue est pour la référence 'mike'. Le mot a donc été reconnu correctement).

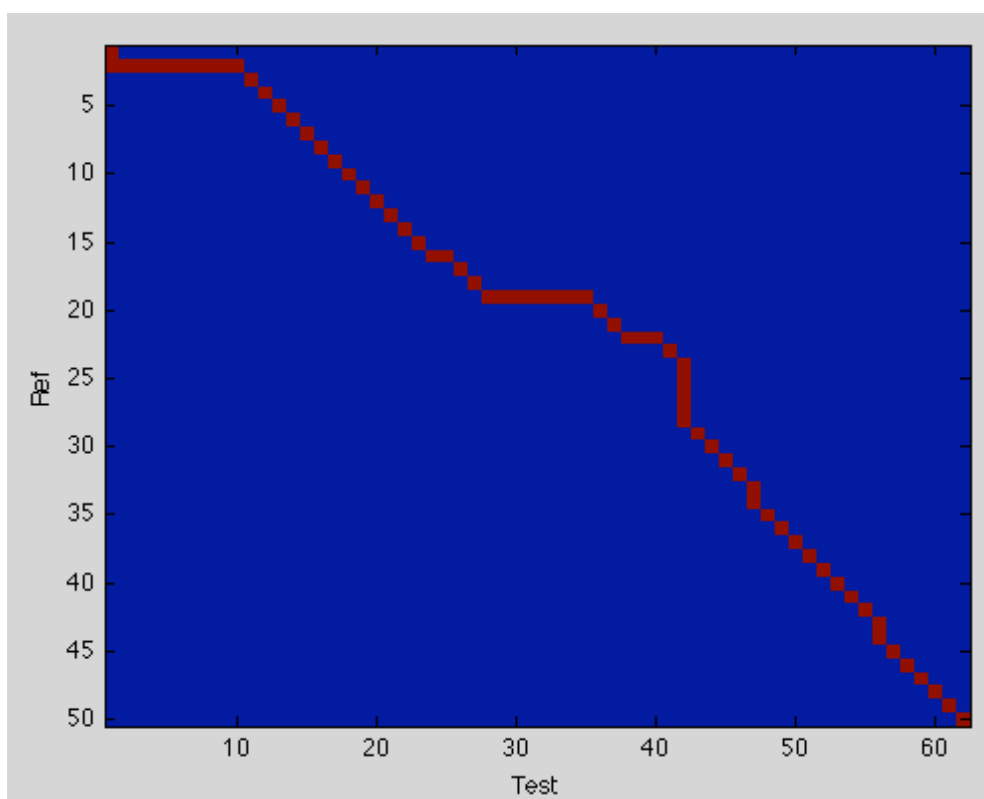


Il vous est demandé d'écrire les fonctions ***dtw_IAN()*** et ***decision()*** .

La fonction ***dtw_IAN()*** :

- Possède deux arguments en entrée : la matrice de coefficients cepstraux du signal à reconnaître et la matrice de coefficients cepstraux d'un signal de référence.
- Renvoie la matrice à deux dimensions des coûts, le coût total et la matrice à deux dimensions contenant le chemin optimal (on pourra matérialiser par la valeur 2 les cellules 'sous' le chemin optimal et par la valeur 1 les autres cellules).
- Vous utiliserez la distance euclidienne entre les couples de vecteurs de paramètres acoustiques. Le développement de cette fonction pourra s'appuyer sur ***l'algorithme donné en TD***.
- La visualisation du chemin pourra se faire par l'appel suivant : `figure, imagesc(chemin), xlabel('Test'), ylabel('Ref')`

Vous devriez obtenir des figures de ce style :



- La fonction ***decision()***, prend en entrée la liste des distances calculées pour l'ensemble du dictionnaire et la liste de mot du dictionnaire et renvoie la distance minimale et le mot associé.