

TDE - Description et analyse du signal sonore

I - Pour débiter ...

1- taille en mémoire :

f_e = fréquence d'échantillonnage $\Rightarrow 44\,100$ points par seconde

γ = 16 bits $\Rightarrow 2$ octets

stéréo $\Rightarrow 2$ pistes (x2)

1 heure $\Rightarrow 3600$ s

Taille = $44\,100 \times 3600 \times 2 \times 2 = 635\,040\,000$ octets

(on divise par 1024×1024) ≈ 606 Mio

2- stéréo $\Rightarrow 2$ pistes

$f_e = 16$ KHz

Taille fenêtrée d'analyse = 256 points (avec recouvrement moitié)

a) 16 000 point / s

b) durée = $\frac{256}{16\,000} = 0,016$ s = 16 ms

c) fichier de 10 min

$$nb = \underbrace{(10 \times 60 \times 16\,000)}_{\text{nb points par 10 min}} / 256 \times 2 - 1 = 74\,999 \text{ fenêtrées}$$

Comme le prof n'a pas précisé si stéréo ou pas dans la question, alors une deuxième réponse peut être donnée :

$nb = 74\,999 \times 2$
 \rightarrow car 2 pistes (stéréo).

II - Energie à court terme: (diapo 32 ou 40)

```
def energie (signal, taille_fenetre):
```

```
    % init
```

```
    recouvrement = floor ( taille_fenetre / 2 )
```

```
    nbfen = floor ( len (signal) / recouvrement ) - 1
```

```
    res = np. zeros (nbfen)
```

```
    % calcul
```

```
    for fen in range (nbfen):
```

```
        I debut = fen * recouvrement
```

```
        res [ fen ] = np. sum ( signal [ I debut : I debut + taille_fenetre ] ** 2 ) / taille_fenetre
```

```
    return res
```

III - Spectrogramme :

```
def spectro (signal, taille_fenetre):
```

```
    % init
```

```
    recouvrement = floor ( taille_fenetre / 2 )
```

```
    nbfen = floor ( len (signal) / recouvrement ) - 1
```

```
    res = np. zeros (1, recouvrement, nbfen)
```

```
    % calcul
```

```
    for fen in range (nbfen):
```

```
        I debut = fen * recouvrement
```

```
        spectre = abs ( np.fft. fft (signal [ I debut : I debut + taille_fenetre ] ) )  
                  calcul spectre
```

res [: fen] = ~~spectre~~ [: floor (len (~~spectre~~) / 2)] . T

return res

ranger 1/2 ~~spectre~~

IV - F0

1- Par analyse temporelle du signal:

d) $F_e = 16 \text{ KHz} \Rightarrow 16000 \text{ Hz}$

nb de points entre 2 pics ≈ 120 (lu sur le graphe de signal)

$$\hookrightarrow \frac{120}{16000} = T_0$$

$$F_0 = \frac{1}{T_0} = \frac{1}{\frac{120}{16000}} = \frac{16000}{120} \approx 133 \text{ Hz}$$

e) plutôt = à priori = plus de chance

\Rightarrow il y a plus de chance que ce soit un homme car pour l'homme c'est entre 50 - 150 Hz tandis que les femmes entre 150 et 300 Hz

2- Par analyse cepstrale:

$$S_n = e_n \otimes C_n$$

\triangleright produit de convolution

a) $S_n = e_n \otimes C_n$

\downarrow FFT (transformé de Fourier)

Domaine temporel

$$\hat{S}_n = \hat{e}_n \times \hat{C}_n$$

\downarrow log

Domaine fréquentiel Spectre

$$\log \hat{S}_n = \log \hat{e}_n + \log \hat{C}_n$$

$$\hat{S}_n = \hat{\hat{e}}_n + \hat{\hat{C}}_n$$

\downarrow FFT⁻¹ (transformé inverse)

Domaine "temporel" Cepstre

b) spectre \rightarrow [spɛktɹ]

cépère \rightarrow [sɛpsɛ]

c) f_0 ?

$$n_0 = 125$$

$$f_e = 16 \text{ kHz} = 16000 \text{ Hz}$$

$$T_0 = \frac{125}{16000}$$

$$f_0 = \frac{1}{T_0} = \frac{1}{\frac{125}{16000}} = \frac{16000}{125} = 128 \text{ Hz}$$

d) Homme

e) $f_0 = 450 \text{ Hz}$

$$f_0 = \frac{1}{T_0} \Rightarrow T_0 = \frac{1}{f_0}$$

$$\frac{n_0}{16000} = \frac{1}{f_0} \Rightarrow n_0 = \frac{16000}{f_0} = \frac{16000}{450} \approx 36$$

Conclusion:

Dans un système de reconnaissance automatique

\rightarrow trentaine de MFCC

\rightarrow 12 MFCC + Δ + $\Delta\Delta \Rightarrow 39$ valeurs
+ énergie
 \rightarrow pas 6 MFCC