

Université Paul Sabatier Toulouse

Compte Rendu

TP 3 : Analyse, codage et synthèse de la parole

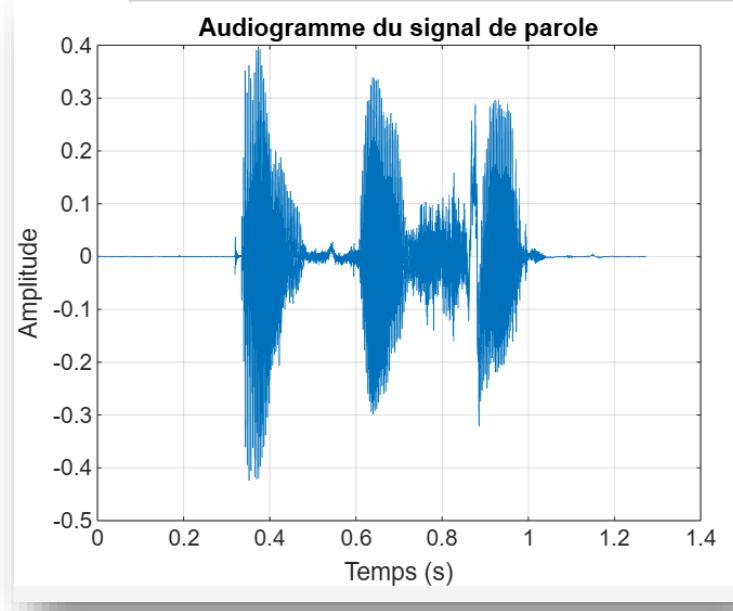
M1 EEA

- ✗ Module : Analyse spectrale des signaux et systèmes
- ✗ Réalisé par :
 - KABOU Abdeldjalil

TP3 : Analyse, codage et synthèse de la parole

1) Analyse :

1.



Les phonèmes correspondent aux parties du signal où l'amplitude et la forme du signal changent.

- ✖ Les voyelles apparaissent comme des segments plus réguliers avec des oscillations bien définies.
- ✖ Les consonnes apparaissent comme des pics brusques ou des zones silencieuses suivies d'une montée rapide.

En supprimant les zones d'activité séparées par des silences au début et à la fin, on obtient un nouveau signal x centré uniquement sur la parole. L'affichage de ce signal x met en évidence cinq blocs principaux correspondant aux cinq phonèmes du mot assécher. Un zoom visuel sur chaque bloc permet d'observer leur forme et de déterminer si le phonème est voisé lorsqu'il présente une structure régulière ou non voisé lorsqu'il s'agit d'un bruit désorganisé.

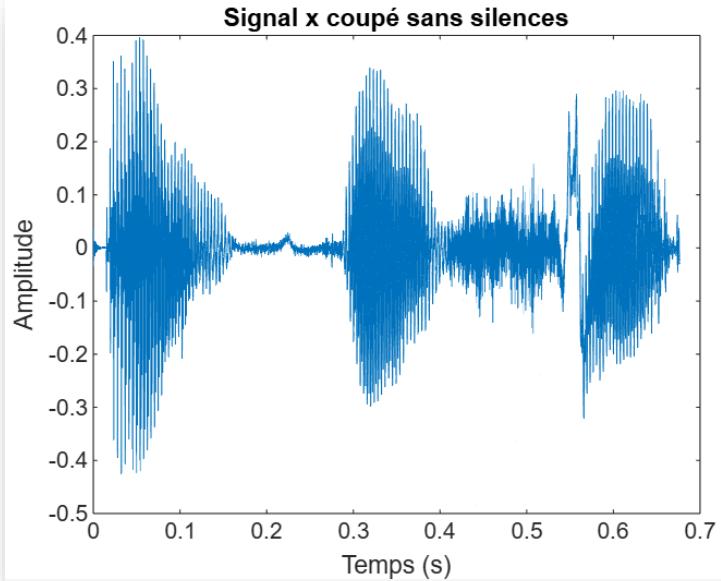
[0.00 – 0.13] → Voisé

[0.13 – 0.26] → Non voisé

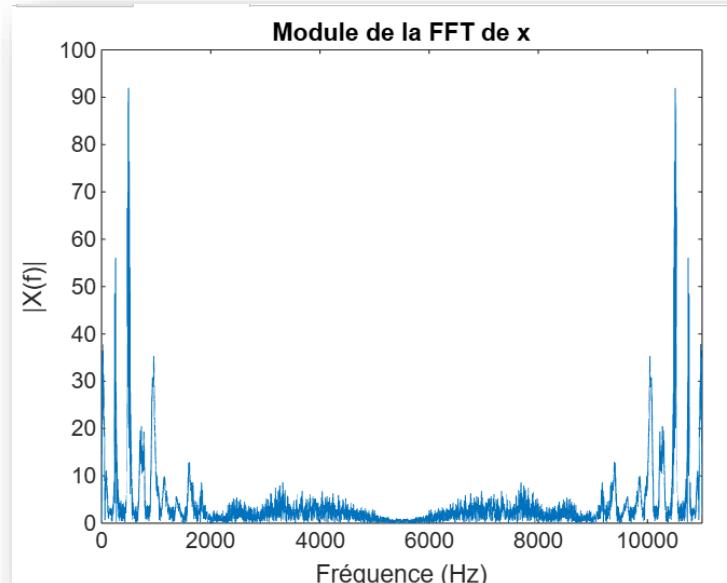
[0.26 – 0.40] → Voisé

[0.40 – 0.57] → Non voisé

[0.57 – 0.70] → Voisé



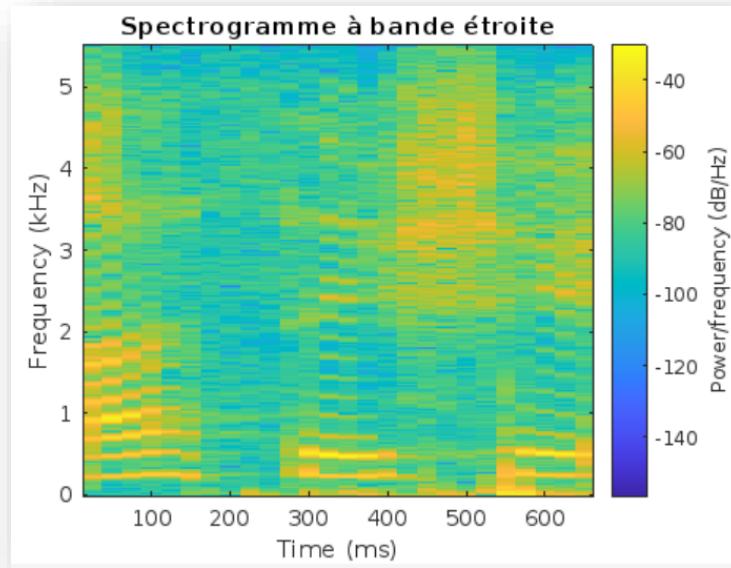
2.



Commentaires :

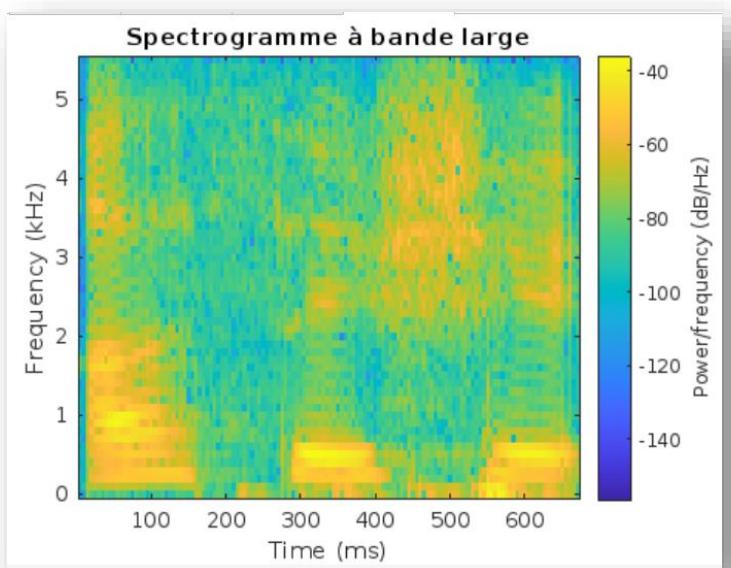
L'analyse du module de la transformée de Fourier du signal x révèle une concentration d'énergie dans les basses fréquences, accompagnée de plusieurs pics harmoniques distincts, indicateurs des sons voisés. Par ailleurs, certaines régions du spectre apparaissent plus diffuses, ce qui est caractéristique des sons non voisés.

3.



Le spectrogramme à bande étroite met en évidence la structure harmonique du signal de parole. L'espacement vertical entre les bandes permet d'estimer la période de pitch, une information précieuse pour l'analyse des phonèmes voisés.

4.



Le spectrogramme à bande large du signal met en évidence cinq segments distincts dans le temps, correspondant aux unités sonores du mot analysé. Chaque segment se caractérise par une structure spectrale unique, traduisant des propriétés acoustiques spécifiques.

Notamment, le troisième et le cinquième segments affichent une distribution d'énergie similaire dans les basses fréquences et une évolution temporelle proche, suggérant une forte proximité acoustique entre eux.

Ainsi, le spectrogramme à bande large permet non seulement d'identifier les différentes composantes du mot, mais aussi de révéler des similitudes spectrales entre certaines d'entre elles.

II. Codage LPC

1.

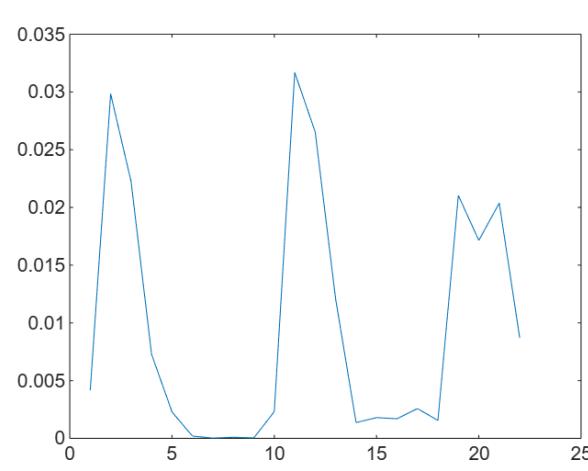
```
Puissance moyenne des tranches :
Columns 1 through 12

 0.0042    0.0298    0.0222    0.0073    0.0023    0.0002    0.0000    0.0001    0.0000    0.0023    0.0317    0.0265

Columns 13 through 22

 0.0121    0.0014    0.0018    0.0017    0.0026    0.0016    0.0210    0.0171    0.0204    0.0087

>>
```



2.

Vecteur des périodes pitch (en secondes) :

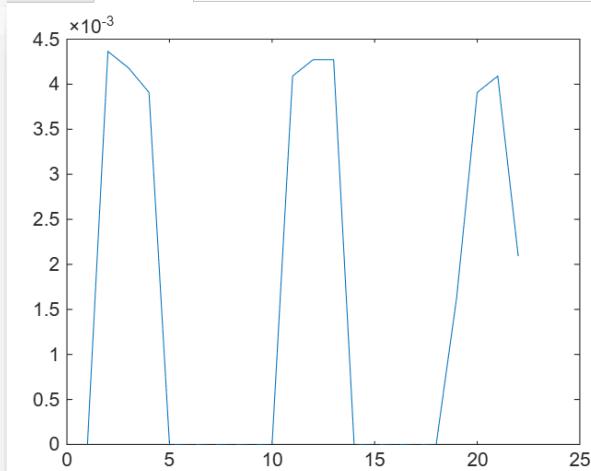
Columns 1 through 12

0.0045	0	0.0042	0	0.0039	0.0042	0.0042	0.0017	0.0025	0.0037	0.0041	0.0043
--------	---	--------	---	--------	--------	--------	--------	--------	--------	--------	--------

Columns 13 through 22

0.0043	0.0040	0.0016	0.0035	0.0016	0.0016	0.0016	0.0039	0.0041	0.0021
--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

>>



3.

7)

$$\alpha(n) + a_1 \alpha(n-1) + a_2 \alpha(n-2) + \dots + a_p \alpha(n-p) = e(n)$$

$\downarrow z$

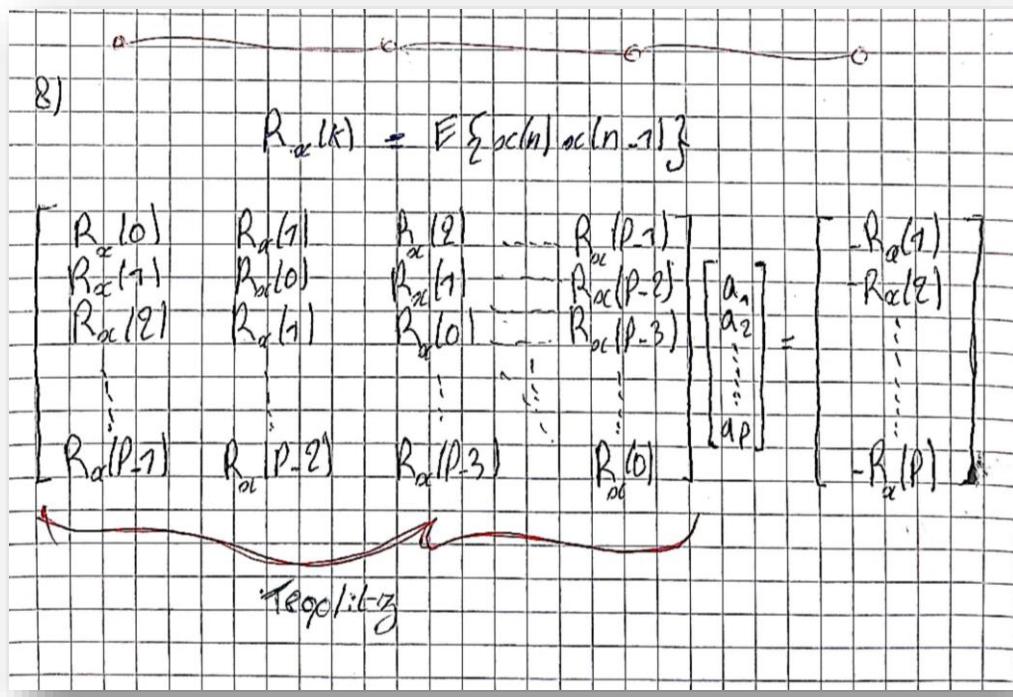
$$X(z) + a_1 X(z) z^{-1} + a_2 X(z) z^{-2} + \dots + a_p X(z) z^{-p} = E(z)$$

$$X(z) (1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_p z^{-p}) = E(z)$$

$$H(z) = \frac{X(z)}{E(z)}$$

$$= \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_p z^{-p}}$$

4.



Ce système linéaire peut être noté simplement : $R \cdot a = -r$

5.

Matrice des coefficients LPC :

-0.6916	0.0239	-0.0780	1.0154	-0.4928	-0.1519	-0.0997	0.6665	-0.2192	-0.0949	0.0227	0.0160
-1.1667	0.3304	-0.0166	0.8795	-0.7091	0.0036	-0.0908	0.7856	-0.5090	0.0240	-0.0569	0.0819
-1.6955	0.9717	-0.2588	0.6217	-0.5556	-0.0681	0.1075	0.5545	-0.5893	0.1420	-0.0926	0.1516
-1.6507	0.8417	-0.1229	0.4327	-0.5818	0.1696	-0.0700	0.6745	-0.7958	0.2540	-0.0538	0.0768
-1.2747	0.2981	-0.2906	0.6280	-0.2760	-0.1142	-0.0168	0.6009	-0.6386	0.1942	-0.2180	0.2394
-0.9311	0.0019	-0.1918	0.2230	0.0818	-0.0387	-0.1299	0.2435	-0.3336	0.2695	-0.1363	0.0830
-0.1005	-0.2248	-0.3633	0.1388	0.1148	0.1132	0.1118	-0.0012	-0.1664	-0.0353	0.0248	0.0130
-0.0590	-0.3054	-0.3736	0.0793	-0.0591	0.2180	0.0145	-0.1650	-0.2237	0.0255	-0.0731	-0.0535
0.1518	0.0291	-0.5062	-0.0912	-0.0334	0.1263	-0.0712	0.0611	-0.3565	-0.0989	-0.1234	0.0204
-1.1021	0.1020	-0.3148	0.3327	-0.1028	0.2614	-0.0701	0.4706	-0.4825	-0.1891	0.1918	0.0118
-1.1677	0.2414	-0.3007	0.1773	-0.0448	0.4194	-0.0835	0.0548	-0.2241	-0.0778	0.3703	-0.2228
-1.0764	0.3048	-0.4434	0.0380	0.1444	0.2960	-0.0595	0.1818	-0.2483	-0.1163	0.3946	-0.2295
-1.0383	0.1447	-0.6249	0.3679	0.0920	0.3683	-0.1041	0.1000	-0.1747	-0.0411	0.0146	0.0021
-0.3080	-0.0291	-0.8436	-0.3096	-0.0451	0.5621	0.3443	0.1782	-0.1857	-0.1815	-0.1791	0.2373
0.5672	0.6171	-0.4413	-0.4262	-0.7236	-0.1149	-0.1010	0.2500	-0.0101	0.0533	-0.1932	0.1252
0.7555	0.7803	-0.2706	-0.4784	-0.9326	-0.3877	-0.2953	0.1177	-0.0065	-0.0412	-0.1560	0.0408
1.0444	0.8843	-0.3689	-0.8387	-1.1446	-0.4178	-0.0012	0.4136	0.1501	-0.0736	-0.2620	-0.1211
0.6839	0.3989	-0.5676	-0.6835	-0.8079	-0.1825	0.0407	0.4200	0.1889	-0.0734	-0.2521	0.0273
-1.0282	0.1878	-0.6256	0.2581	0.0857	0.2907	-0.0148	-0.0502	0.0367	-0.2086	0.1347	-0.0533
-1.0971	0.2721	-0.5891	0.3053	0.1035	0.2128	-0.0121	-0.0601	0.1687	-0.2067	0.0405	-0.0597
-0.9396	0.2751	-0.5669	0.1906	-0.1048	0.5094	0.0461	-0.0832	-0.0515	-0.0035	0.0530	-0.0319
-0.6166	0.2156	-0.9116	0.0210	-0.1146	0.7805	0.1408	0.1708	-0.2383	-0.1881	-0.0891	0.1724

>>

6.

Calcul des bits par tranche :

- ✗ Puissance moyenne : 8 bits
- ✗ Période pitch : 8 bits
- ✗ Coefficients LPC : $12 \times 8 = 96$ bits
- ✗ Total par tranche : $8 + 8 + 96 = 112$ bits

Total des bits pour l'ensemble du signal :

Si le signal est segmenté en nbtranches tranches, le total de bits nécessaires est :

$$\text{Total bits signal entier} = \text{nbtranches} \times 112$$

Comparaison avec un codage direct du signal :

En supposant un encodage de chaque échantillon sur 8 bits, le signal entier occuperait :

$$\text{Total bits sans codage} = N \times 8$$

Conclusion :

Le codage basé sur la puissance moyenne, la période de pitch et les coefficients LPC permet de réduire significativement le nombre de bits nécessaires par rapport à une transmission brute du signal. Cette compression optimise le stockage et la transmission tout en préservant les informations essentielles à la reconstruction du signal de parole.

III. Synthèse

- I. En multipliant la période de pitch par deux, le signal synthétisé devient notablement plus grave. Cette modification résulte du fait que la fréquence fondamentale des segments voisés est divisée par deux, entraînant une baisse de la hauteur du son. En conséquence, les phonèmes voisés paraissent ralentis et plus profonds, donnant l'impression d'une voix plus grave ou plus masculine. En revanche, les phonèmes non voisés, générés par un bruit blanc, restent inchangés, car leur excitation ne dépend pas de la période de pitch. Cette expérience illustre bien l'impact de la fréquence d'excitation sur la perception du timbre vocal et trouve des applications dans la synthèse et la transformation de la parole.

- II. En fixant une période de pitch constante, le signal synthétisé devient monotone et artificiel. La voix perd ses variations naturelles, rendant la parole moins expressive et plus robotique. Cette expérience montre que le pitch est essentiel pour donner du naturel et de l'intonation à la parole.

III. Réponse en fréquence du modèle AR :

On obtient la réponse en fréquence en évaluant $H(z)$ sur le cercle unité $z=\exp(j\omega)$:

$$H(e^{j\omega}) = \frac{1}{1 + a_1e^{-j\omega} + a_2e^{-j2\omega} + \cdots + a_Pe^{-jP\omega}}$$

Avec Matlab, on peut utiliser la fonction freqz pour calculer et tracer la réponse en fréquence.

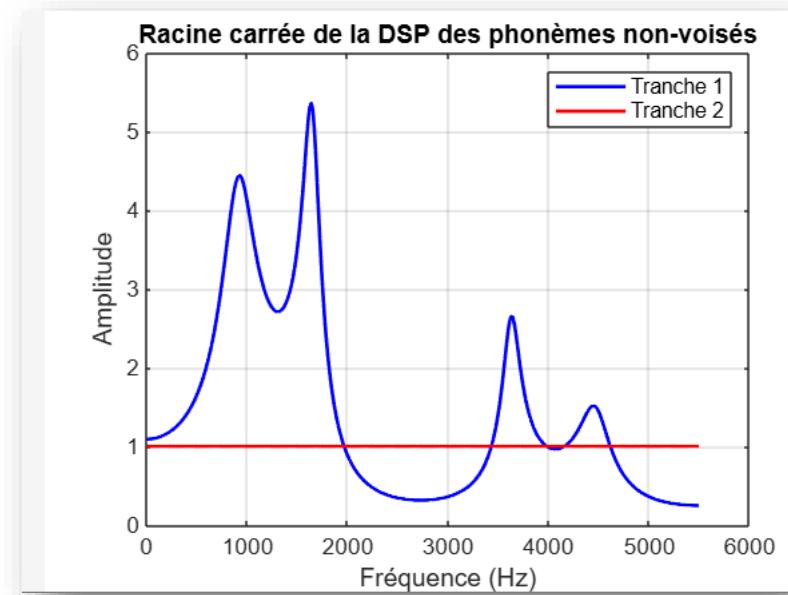
DSP du signal $x(n)$ pour un bruit blanc $e(n)$:

Si $e(n)$ est un bruit blanc de puissance σ_e^2 , alors la DSP de $x(n)$, notée $S_x(\omega)$, est donnée par :

$$S_x(\omega) = |H(e^{j\omega})|^2 \cdot \sigma_e^2$$

La racine carrée de la DSP est donc :

$$\sqrt{S_x(\omega)} = |H(e^{j\omega})| \cdot \sqrt{\sigma_e^2}$$



La racine carrée de la DSP montre comment l'énergie spectrale est distribuée dans les tranches non-voisées. Ces segments présentent généralement une distribution plus étalée en fréquence,

ce qui est caractéristique des sons non-voisés qui ressemblent à du bruit. En observant ces courbes, on peut mieux comprendre la structure spectrale du signal de parole et son comportement en présence d'un bruit blanc d'excitation.

IV. Améliorations

Pour améliorer la qualité du signal synthétisé, on peut utiliser des fenêtres de Hanning de 30 millisecondes avec un recouvrement de 50% entre les fenêtres successives. Cela améliore la continuité entre les tranches en réduisant les discontinuités aux bords des fenêtres.