

Projet : Vocodeur de phase

Introduction

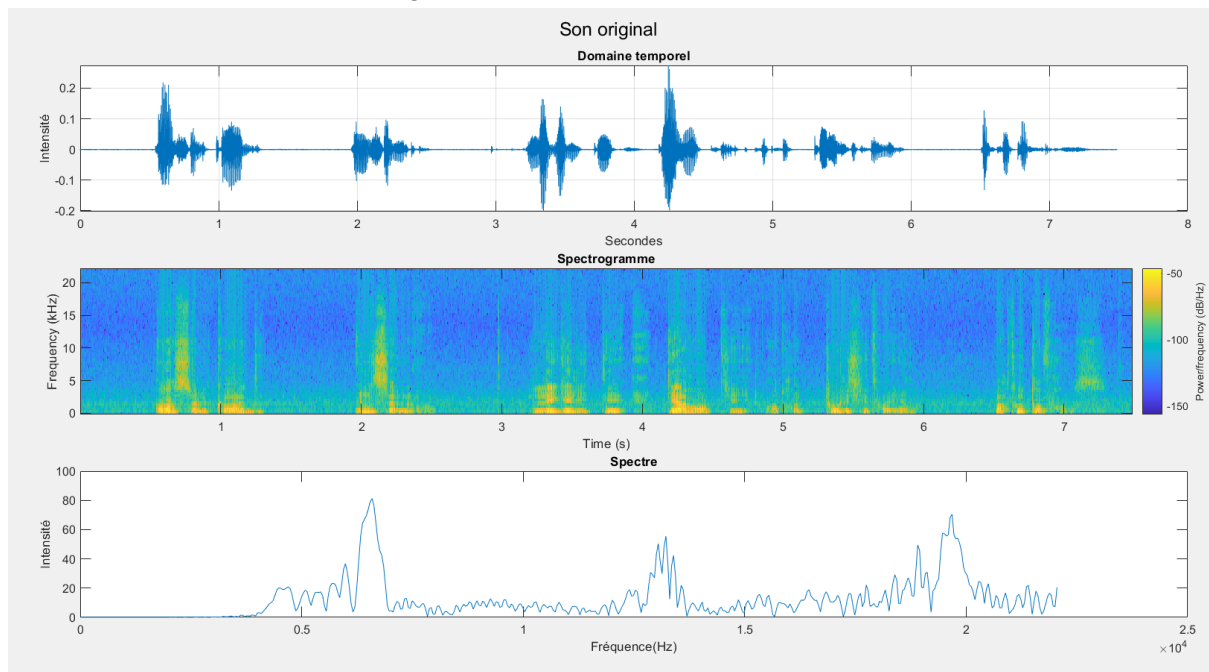
Ce projet vise à créer un vocodeur de phase. L'objectif est donc de rendre "robotiques" des audios, principalement des paroles.

Nous avons 3 audios provenant de divers médias et 1 audio que nous avons enregistré nous-mêmes et appelé "Civil.wav". Nous avons utilisé les 4 audios et avons choisi celui que nous avons produit pour illustrer les effets de nos traitements.

Partie 1

Après avoir enregistré notre fichier audio, nous avons visualisé ses différentes caractéristiques.

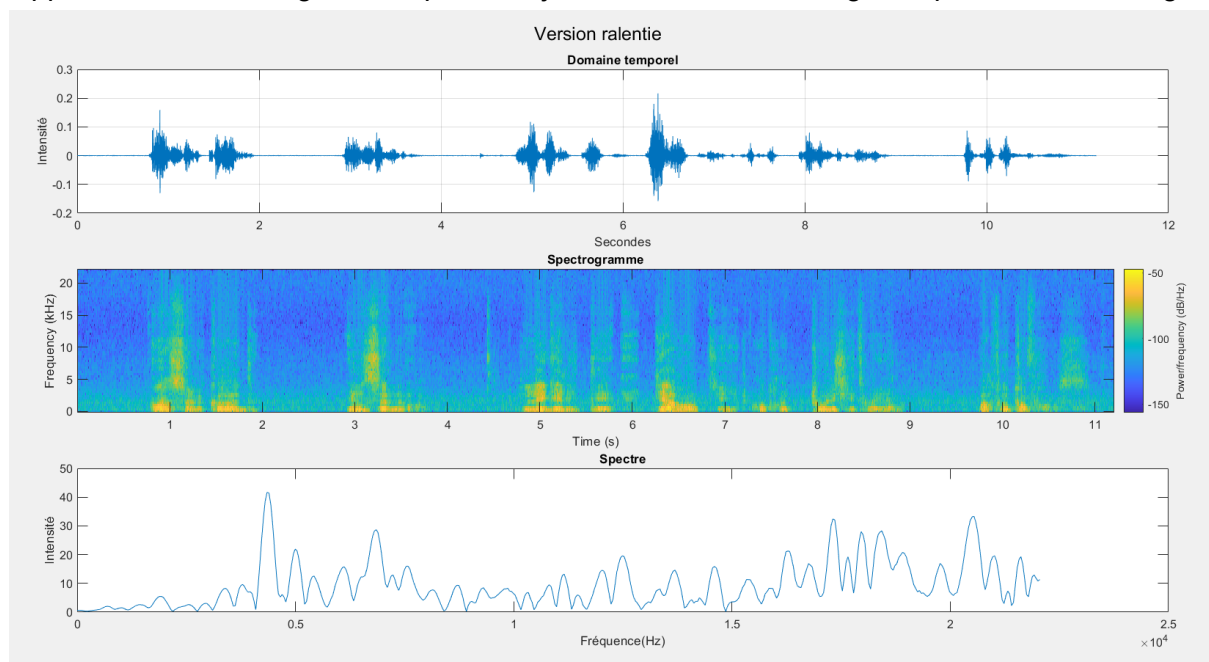
Voici les courbes correspondant au signal "Civil.wav" non modifié : son allure dans le domaine temporel, son spectrogramme et son spectre.



Nous avons joué sur la vitesse de lecture de l'audio : Notre première expérience fut de ralentir sa lecture sans en changer le pitch. Nous avons décidé de le lire aux $\frac{2}{3}$ de sa vitesse initiale.

Pour ce faire, nous utilisons plusieurs fonctions de modification.

Premièrement, nous allons modifier le signal qui est, dans notre cas, de la parole, en créant des transformées de Fourier à court terme dans des fenêtres temporelles relativement courtes (de 20 à 30 ms). En conséquence, on peut considérer le signal comme stationnaire pour chaque fenêtre; à la suite de quoi nous modifions chaque portion du signal (chaque fenêtre) grâce à une interpolation de la TFCT. Cela permet de moduler la taille de chaque fenêtre. Une fois cela fait, on revient dans le domaine temporel. La modification de la taille de la fenêtre aura permis de ralentir (comme dans le cas présent) ou d'accélérer l'audio. On ne modifie pas les fréquences en y rajoutant ou en enlevant des valeurs, mais en rapprochant ou en éloignant les points déjà existants dans des signaux plus ou moins longs.

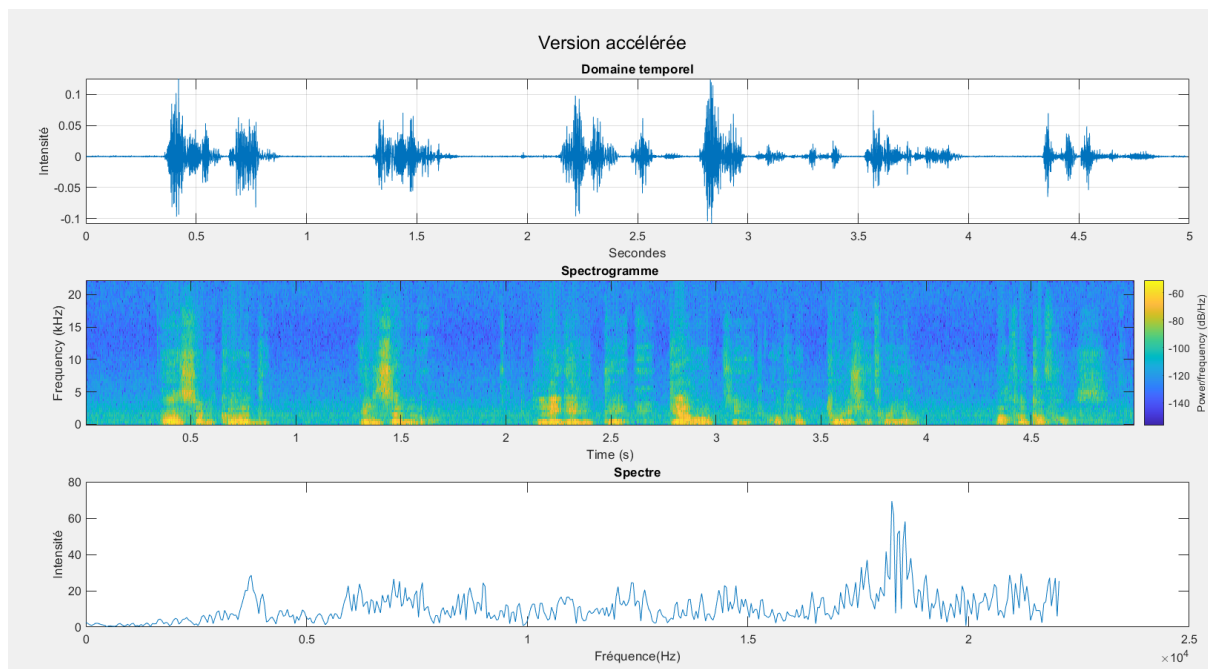


Tout d'abord, nous pouvons noter que la taille de la fenêtre passe bien de 8 à 12 secondes en raison d'une augmentation proportionnelle de la durée du signal. Il y a également une augmentation de l'intensité sonore générale du signal.

Le spectrogramme reste tout à fait similaire, à cela près qu'il couvre environ 12 secondes au lieu de 8, ce qui est normal vu que l'on conserve le même pitch.

Nous pouvons constater que le spectre est plus "mouvementé", mais que son intensité est grossièrement divisée par deux. Il est normal d'avoir une intensité plus faible en raison de la durée augmentée du signal. Les pics d'intensité de l'original sont répartis sur une plus longue durée, donc moins intenses. Les fréquences plus basses (resp. hautes) ont une intensité plus élevée (resp. faible) par rapport au signal d'origine.

Voici la version accélérée du signal. Le principe est le même que pour la version ralentie, le rapport ici est simplement de 3/2.



Nous observons que la durée du signal est bien de 5 secondes, la courbe du domaine temporel est bien la même que celle d'origine. Il y a également une diminution de l'intensité sonore du signal.

Le spectrogramme reste le même, sur 5 secondes au lieu de 8.

Dans le spectre, les hautes (resp. basses) fréquences ont augmenté (resp. baissé) en intensité, les pics sont plus resserrés. Il est intéressant de constater que l'échelle de l'intensité (en ordonnée) n'atteint pas les 100 comme pour le signal original. On observe que l'intensité est plus répartie sur tout le spectre des fréquences par rapport à l'original, cela est peut-être dû à une dégradation de la qualité due au traitement (le signal semble plus numérique). Le fait que les pics les plus hauts restent en-dessous des plus hauts de l'original peut être expliqué par une chute plus rapide, empêchant la confusion et l'addition de deux pics autrement distincts.

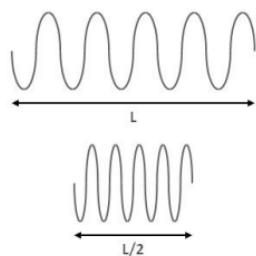


Illustration du phénomène de contraction du signal :

Si l'on prend un signal A, que l'on le contracte pour le faire "rentrer" dans une fenêtre plus courte, le son du signal sera automatiquement plus aigu. C'est pourquoi, lors de la contraction, il faut faire attention à garder la même fréquence fondamentale. Cela peut occasionner des effets "robotiques".

Partie 2

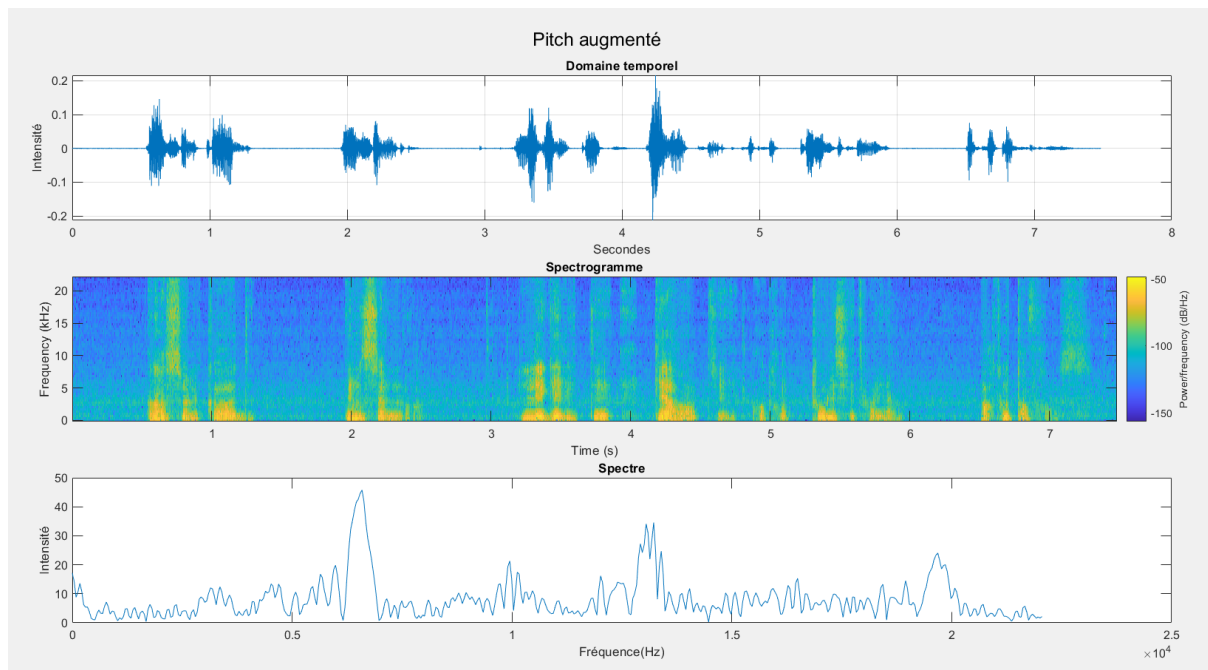
Après le succès retentissant et éclatant de la modification temporelle du signal (en ayant réussi à ne pas toucher au pitch), vos serviteurs ont décidé de tester à nouveau leurs compétences en faisant cette fois-ci une modification du pitch sans pour autant modifier la vitesse de lecture.

Pour ce faire, nous utilisons plusieurs fonctions de modification.

Nous allons commencer par modifier le signal en créant des TFCT dans des fenêtres similaires à celles de la première partie. Puis nous effectuons l'interpolation (comme dans la partie précédente). Le ré-échantillonnage (via la fonction matlab resample) permet de moduler le nombre de points présents dans une fenêtre, tout en conservant la même période d'échantillonnage. Une fois cela fait, on revient dans le domaine temporel. La modification du nombre de points présents dans les fenêtres aura permis d'augmenter (comme dans le cas présent) ou de réduire le pitch.

En résumé, on augmente (resp. réduit), selon un ratio défini, le nombre de points, augmentant (resp. réduisant) ainsi la taille de chaque fenêtre, puis on réduit (resp. augmente) la taille de chaque fenêtre par l'inverse de ce ratio, afin de retrouver un signal de même durée que le signal initial (et la même période d'échantillonnage), mais avec un pitch plus haut (resp. bas).

Visualisation des courbes du signal à pitch augmenté



Dans le domaine temporel, nous avons un signal visuellement semblable à l'original. L'intensité sonore l'est également.

Le spectrogramme a un aspect bruité et numérique. Les fréquences sont plus intenses.

Le spectre montre des pics d'une intensité plus faible que l'original, ainsi que l'apparition de fréquences basses ($< 0,5 \times 10^4$ Hz).

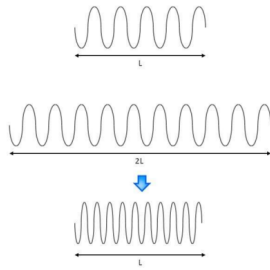
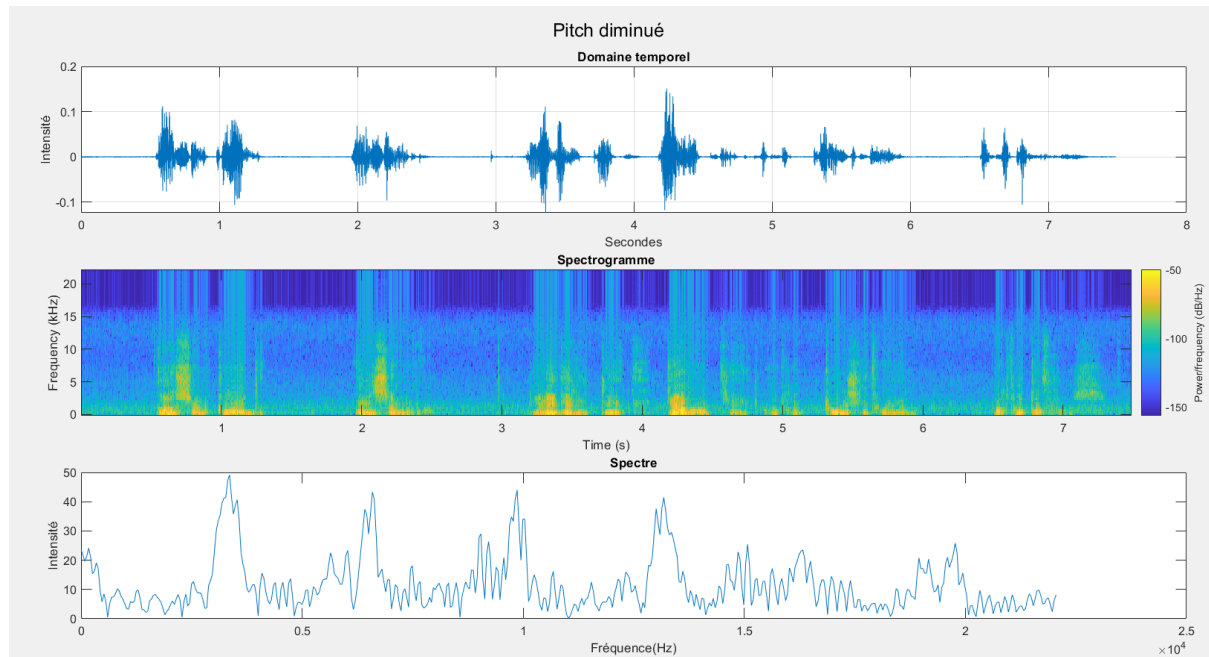


Illustration de l'augmentation du pitch par interpolation

Afin d'augmenter le pitch, l'on procède dans un premier temps à un doublement du nombre de points et de la fenêtre du signal (comme si on avait additionné le signal à lui-même). Ensuite, on prend le signal résultant pour le contracter dans une fenêtre de la longueur du signal original. Cela résulte en une augmentation du pitch, car les fréquences auront augmenté.

Visualisation des courbes du signal à pitch diminué



Le domaine temporel est toujours semblable à l'original.

Le spectrogramme montre une baisse notable d'intensité des fréquences (surtout les hautes fréquences).

Pour ce qui est du spectre, on peut observer plusieurs nouveaux pics, de plus faible intensité que les pics du signal original. Moins de "vallées", qui sont de plus forte intensité.

Cette entreprise s'est avérée quelque peu complexe. Le principal problème rencontré était que, contrairement à la modification de la vitesse de lecture, il était important ici que les vecteurs représentant les signaux audios soient de taille identique (que le signal modifié et le signal de taille identique soient de même durée). Cela impliquait qu'il fallait reformater le vecteur le plus long vers la taille du vecteur le plus court entre le signal d'origine et le signal ré-échantillonné.

Partie 3

Une fois les écueils de la vitesse de lecture et du pitch surmontés, nous nous attelâmes à la robotisation de la voix.

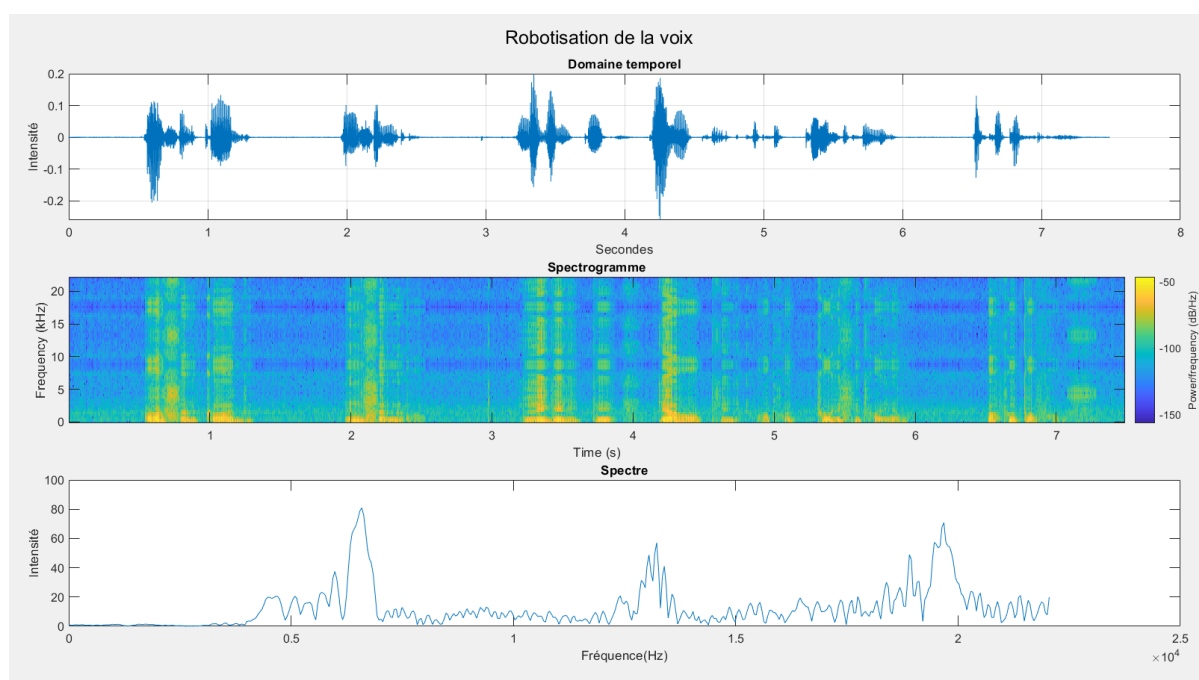
Pour cela, nous utilisâmes une sinusoïde (qui est en réalité une exponentielle complexe, fonction d'une fréquence fondamentale décidée arbitrairement).

Nous créâmes un vecteur vide. En chaque point dudit vecteur, nous affectâmes la valeur correspondante du signal d'origine multipliée par la sinusoïde. Cela créa une légère variation du son perçu par rapport au signal d'origine. Cette variation étant appliquée pour chaque valeur du vecteur, nous obtînâmes un signal légèrement différent du signal d'origine, mais pas de façon très significative. Pour pallier ce problème, nous implémentâmes un intervalle, permettant de conserver sur un certain nombre de points consécutifs du signal (définis par l'utilisateur du programme) le même son modifié par la sinusoïde. En somme, ce fut une discrétisation modulable du signal.

Choisir une fréquence originale de 200, 500, 1000 ou 2000 hz n'a aucune incidence sur le résultat final audio.

Paramètres : fréquence fondamentale : 500 hz

Intervalle : 5



Dans le domaine temporel, nous observons une symétrie au signal original par rapport à l'axe des abscisses : l'intensité est inversée.

Sur le spectrogramme, nous pouvons observer clairement les moments de parole et les moments de silence. Il y a un bruit constant sur les fréquences faibles (bruit de fond venant du signal original). Les sons étant divisés en intervalles, il est logique de voir se maintenir dans le temps les fréquences émulées.

Le spectre est totalement identique au spectre original.

Conclusion

Lors de la réalisation de ce projet, nous sommes parvenus à modifier la vitesse de lecture de fichiers audio, modifier leur pitch, leur appliquer un effet de robotisation. Ces opérations de traitement de signal dans le but d'apporter des modifications audible à un extrait audio nous on permis d'accentuer notre compréhension de toutes les petites étapes nécessaire à la création d'un signal audio numérique, notamment à partir d'un signal déjà existant.