Implémentation d'Effets Audio par Traitement du Signal

Karim Sadiki

***Résumé*—Ce projet présente l'implémentation de quatre effets audio classiques : le tremolo/vibrato, la réverbération par convolution, les effets de modulation temporelle (chorus et flanger), ainsi que le pitch shifting. Chaque effet est développé en utilisant uniquement des techniques de traitement du signal numérique. Les algorithmes implémentés incluent la modulation d'amplitude et de fréquence par LFO, une FFT Cooley-Tukey pour la convolution rapide, des lignes de délai modulées, et un vocodeur de phase pour la modification de hauteur tonale. Les résultats sont évalués uniquement sur des signaux vocaux à travers des analyses temporelles et spectrales. Les spectrogrammes comparatifs démontrent l'efficacité des traitements implémentés.**

***Index Terms*—Traitement du signal audio, DSP, tremolo, vibrato, réverbération, convolution, FFT, chorus, flanger, pitch shifting, phase vocoder, formants, STFT, LFO.**

I. Introduction

LES effets audio constituent un domaine fondamental du traitement du signal numérique, avec des applications allant de la production musicale à la synthèse vocale. Historiquement, ces effets sont apparus bien avant l'ère de l'apprentissage automatique et reposent exclusivement sur des opérations DSP classiques : filtrage, modulation, convolution, interpolation et transformations spectrales [1].

Ce projet vise à implémenter quatre catégories d'effets audio parmi les plus utilisés dans l'industrie musicale :

1) **Tremolo et Vibrato** : modulation d'amplitude et de fréquence par oscillateur basse fréquence (LFO).
2) **Réverbération par convolution** : simulation de l'acoustique d'un espace par convolution avec une réponse impulsionnelle synthétique.
3) **Chorus et Flanger** : effets de modulation temporelle basés sur des lignes de délai variables.
4) **Pitch Shifting avec Formant Shifting** : modification de la hauteur tonale avec préservation optionnelle des caractéristiques vocales.

L'objectif pédagogique principal est de renforcer la compréhension pratique des liens entre les opérations mathématiques du traitement du signal et les phénomènes auditifs perçus. Chaque effet est implémenté *from scratch*, c'est-à-dire sans appel à des fonctions issues de librairies telles que Librosa.

Cet article est organisé comme suit : la Section II présente les principes théoriques et les équations fondamentales de chaque effet. La Section III expose les résultats expérimentaux. Enfin, la Section IV conclut l'article.

II. Méthodologie et Fondements Théoriques

A. Tremolo et Vibrato

Le **tremolo** et le **vibrato** sont deux effets de modulation fondamentaux, souvent confondus mais fondamentalement différents [1] : le tremolo est une modulation périodique de l'*amplitude* du signal, tandis que le vibrato est une modulation périodique de la *fréquence* (hauteur) du signal.

Le tremolo applique une modulation d'amplitude via un LFO (Low Frequency Oscillator) :

$$y_{trem}(t) = x(t)\cdot g(t) \quad \text{où} \quad g(t) = (1-d)+d\cdot\frac{\sin(2\pi f_{LFO}\cdot t)+1}{2} \tag{1}$$

avec $d \in [0,1]$ la profondeur de modulation et f_{LFO} la fréquence du LFO (typiquement 4–8 Hz). Le gain $g(t)$ varie entre $(1-d)$ et 1.

Le vibrato module le temps de lecture du signal, créant une variation de fréquence perçue :

$$y_{vib}(t) = x(t+\tau(t)) \quad \text{où} \quad \tau(t) = d_{vib}\cdot\sin(2\pi f_{vib}\cdot t) \tag{2}$$

avec d_{vib} la profondeur en secondes (typiquement 1–5 ms) et f_{vib} la fréquence de modulation (4–7 Hz).

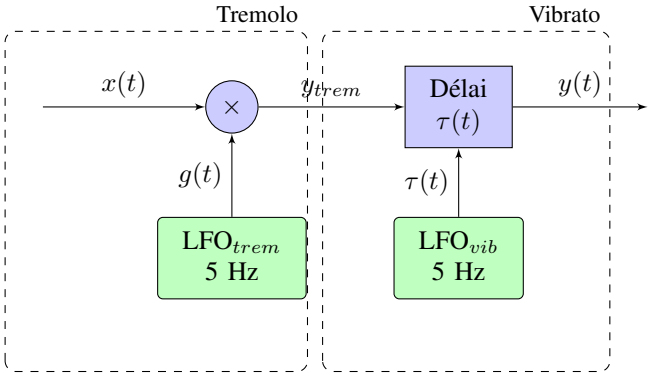

FIGURE 1 – Schéma bloc combiné Tremolo + Vibrato en cascade

B. Réverbération par convolution

La réverbération naturelle résulte des multiples réflexions du son sur les surfaces d'un espace acoustique. Mathématiquement, elle peut être modélisée comme la convolution du signal source $x[n]$ avec la réponse impulsionnelle (IR) de la salle $h[n]$ [2] :

$$y[n] = (x * h)[n] = \sum_{k=0}^{L-1} x[n-k]\cdot h[k] \tag{3}$$

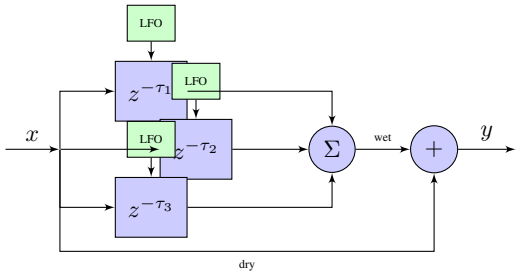

FIGURE 3 – Schéma bloc de l'effet Chorus à 3 voix.

où L est la longueur de la réponse impulsionnelle.

Notre implémentation génère une IR synthétique composée de trois éléments [3] : (1) l'impulsion directe et pré-délai $h_{direct}[n] = \delta[n - n_{predelay}]$, (2) les réflexions précoces $h_{early}[n] = \sum_{i=1}^{N_{ref}} a_i \cdot \delta[n - n_i]$, et (3) la queue diffuse avec décroissance RT60 :

$$h_{tail}[n] = \mathcal{N}(0,1) \cdot e^{-\frac{6.91 \cdot n}{RT60 \cdot f_s}} \tag{4}$$

Le théorème de convolution permet d'utiliser la FFT pour réduire la complexité de $O(N \cdot L)$ à $O(N \log N)$ [4] : $y = \mathcal{F}^{-1}\{\mathcal{F}\{x\} \cdot \mathcal{F}\{h\}\}$.

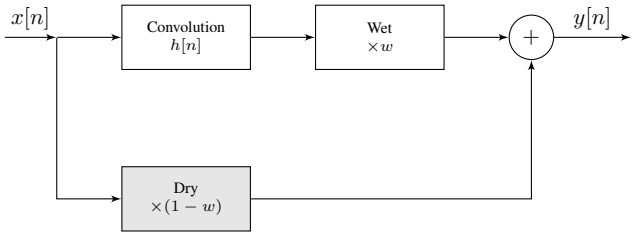

FIGURE 2 – Schéma bloc simplifié de la réverbération par convolution

C. Chorus et Flanger

Le chorus et le flanger sont des effets de modulation temporelle basés sur des lignes de délai dont le temps de retard varie périodiquement [5]. Le signal de sortie est :

$$y(t) = (1 - mix) \cdot x(t) + \frac{mix}{N} \sum_{i=1}^{N} x(t - \tau_i(t)) \tag{5}$$

où N est le nombre de voix et $\tau_i(t) = \tau_{base,i} + d_{mod} \cdot \sin(2\pi f_{LFO} \cdot t + \phi_i)$ avec $\phi_i = \frac{2\pi i}{N}$ le déphasage entre les voix.

TABLE I – Comparaison des paramètres Chorus et Flanger

Paramètre	**Chorus**	**Flanger**
Délai de base	15–30 ms	1–10 ms
Profondeur	3–10 ms	1–5 ms
Fréquence LFO	0.5–2 Hz	0.1–1 Hz
Feedback	0–0.3	0.3–0.8
Nombre de voix	2–4	1
Effet perçu	Épaisseur	Effet "jet"

Le flanger utilise des délais plus courts, créant un effet de filtre en peigne (comb filter) dont les notches se déplacent dans le spectre.

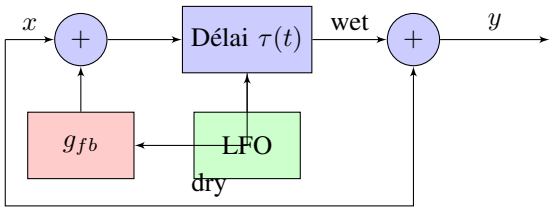

FIGURE 4 – Schéma bloc de l'effet Flanger avec boucle de feedback.

D. Pitch Shifting et Formant Shifting

Le pitch shifting modifie la hauteur tonale sans changer la durée. L'approche utilise le *phase vocoder* [6], [7] : (1) Analyse STFT, (2) Time stretching par facteur α, (3) Resampling pour retrouver la durée originale. Le facteur de transposition en demi-tons s détermine : $\alpha = 2^{s/12}$.

La phase est accumulée pour maintenir la cohérence entre trames :

$$\phi_{out}[m,k] = \phi_{out}[m-1,k] + H_{out} \cdot \omega_k + \Delta\phi \tag{6}$$

Sans correction, le pitch shifting modifie les formants (effet "chipmunk"). La correction utilise l'enveloppe spectrale [8] : $X_{corr}[k] = X_{shifted}[k] \cdot E_{orig}[k]/E_{shifted}[k]$.

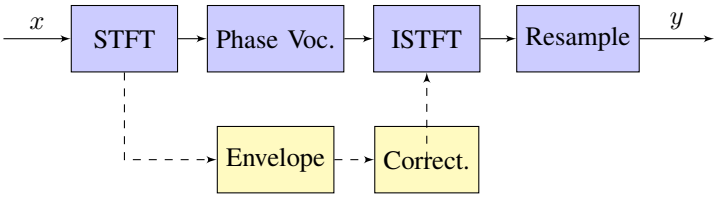

FIGURE 5 – Schéma bloc du Pitch Shifter (chemin pointillé : correction des formants).

III. RÉSULTATS EXPÉRIMENTAUX

A. Signal d'entrée

Les effets ont été appliqués sur un corpus vocal provenant de la base CMU ARCTIC, échantillonné à 16 kHz. Le signal d'entrée sélectionné présente une durée d'environ 7 secondes.

L'analyse spectrale de la Figure 6 montre que la bande de fréquence principale s'étend de 100 Hz à 8 kHz. Les harmoniques vocales sont clairement visibles et espacées régulièrement, créant une structure en peigne caractéristique de la phonation voisée. Le fondamental F0 se situe approximativement entre 120 et 150 Hz, typique d'une voix masculine. Les formants vocaux F1, F2 et F3 sont distincts dans les zones respectives de 500 Hz, 1500 Hz et 2500 Hz. L'énergie spectrale est concentrée principalement sous 4 kHz, et les zones de silence entre les énoncés apparaissent comme des bandes sombres verticales. Aucune présence de bruit basse fréquence ou de distorsion harmonique n'est détectable, confirmant la qualité du signal source.

L'analyse temporelle illustrée par la Figure 7 révèle une enveloppe temporelle claire avec alternance de segments de

parole et de silences. L'amplitude est normalisée entre -1 et +1, présentant une structure rythmique bien définie avec des pics d'intensité correspondant aux consonnes plosives. Le signal présente un rapport signal/bruit élevé sans distorsion visible.

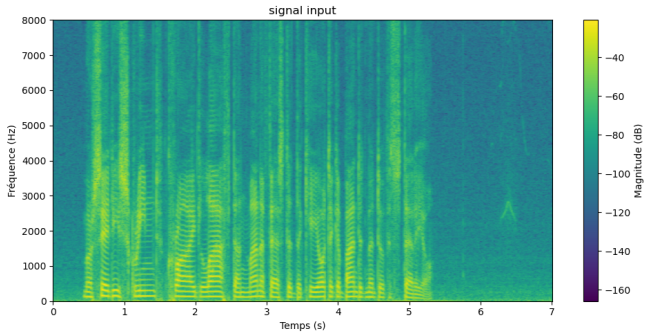

FIGURE 6 – Spectrogramme du signal d'entrée.

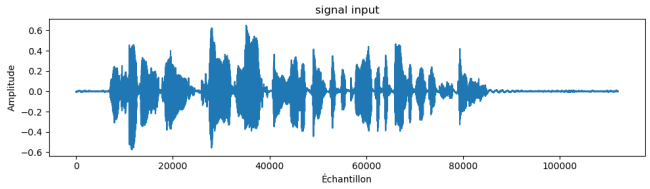

FIGURE 7 – Forme d'onde du signal d'entrée.

B. *Tremolo et Vibrato*

Le spectrogramme de la Figure 8 révèle des transformations caractéristiques dans le domaine fréquentiel. La modulation spectrale induite par le tremolo se manifeste par une variation périodique de l'intensité des harmoniques dans le temps, visible comme des vagues verticales traversant toutes les bandes de fréquence simultanément. Cette modulation affecte uniformément tout le spectre, ce qui est la signature caractéristique d'une multiplication dans le domaine temporel.

La Figure 9 présente des modifications temporelles significatives par rapport au signal d'entrée. L'effet le plus notable est la modulation d'amplitude périodique induite par le tremolo. L'enveloppe globale du signal présente désormais une oscillation visible à 5 Hz, créant un effet de pulsation rythmique caractéristique. Les pics d'amplitude ne sont plus constants mais varient selon une sinusoïde de période approximative de 200 ms, soit l'inverse de la fréquence du LFO. Contrairement au signal original où l'enveloppe suit naturellement l'intensité phonétique, le signal traité montre une modulation artificielle superposée, avec des maxima et minima réguliers même pendant les segments à énergie constante. Malgré cette modulation prononcée, la structure rythmique et les transitions entre phonèmes restent identifiables, confirmant que l'effet n'introduit pas de distorsion non-linéaire.

L'effet du vibrato se distingue par un rétrécissement spectral des harmoniques vocales. Alors que ces dernières apparaissaient comme des lignes horizontales fines dans le signal d'entrée, elles présentent maintenant un rétrécissement horizontal marqué. Ce phénomène résulte directement de la modulation de fréquence, où chaque composante spectrale est étalée sur une plage de ±f autour de sa fréquence centrale. Une analyse fine révèle l'apparition de bandes latérales autour de chaque harmonique, espacées de ±fLFO. Ces bandes sont le résultat mathématique de la modulation de fréquence. Un aspect important est la préservation des formants. Les régions de concentration d'énergie correspondant aux formants F1, F2 et F3 restent localisées aux mêmes fréquences centrales que dans le signal d'entrée, indiquant que le vibrato ne déplace pas la position moyenne des formants mais crée une vibration autour de ces positions. L'évaluation perceptive confirme un effet musical expressif, similaire au vibrato naturel d'un chanteur. La profondeur de 0.7 pour le tremolo produit une modulation notable sans être excessive, tandis que le vibrato à 3 ms de profondeur reste dans la plage naturelle.

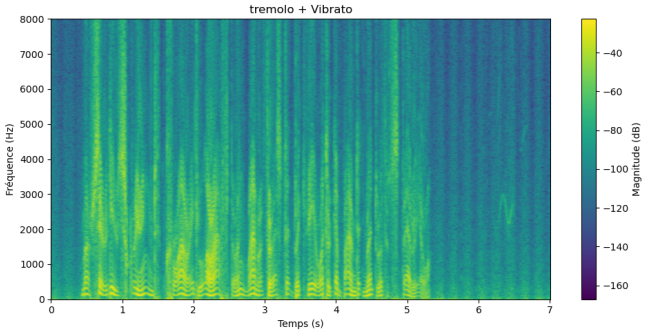

FIGURE 8 – Spectrogramme après application des effets de tremolo et vibrato.

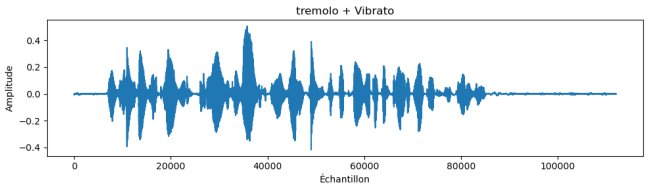

FIGURE 9 – Forme d'onde après application des effets de tremolo et vibrato.

C. *Réverbération par convolution*

La Figure 10 illustre une transformation acoustique profonde du signal temporel. L'extension temporelle constitue la modification la plus immédiatement visible. Alors que le signal d'entrée présente des transitions nettes entre segments de parole et silences, le signal réverbéré montre une persistance du son après chaque événement sonore. Les zones de silence sont maintenant remplies par la queue de réverbération, créant une continuité acoustique absente du

signal original. Le pré-délai de 30 ms est observable comme un décalage entre l'attaque du signal original et l'apparition des réflexions précoces, créant une séparation claire entre le son direct et l'effet spatial. Cette caractéristique est essentielle pour simuler la distance physique entre la source sonore et les premières surfaces réfléchissantes d'une salle. Après chaque segment de parole, l'amplitude ne retombe plus instantanément à zéro mais décroît exponentiellement selon le RT60 configuré à 2.0 s. Cette décroissance suit approximativement la loi A(t) = A × exp(-6.91t/RT60), conforme à la théorie de la réverbération de Schroeder. Les réflexions précoces ajoutent une texture granulaire fine dans les 80 à 100 premières millisecondes, absente du signal d'entrée, correspondant aux premières réflexions sur les parois de la salle virtuelle. Un effet secondaire notable est le masquage temporel, où les consonnes faibles et les transitions rapides sont noyées dans la queue de réverbération des phonèmes précédents, réduisant l'intelligibilité mais augmentant considérablement la sensation d'espace. Le spectrogramme de la Figure 11 révèle des transformations spectrales caractéristiques de l'acoustique des salles. L'extension temporelle des harmoniques est particulièrement frappante. Les harmoniques vocales, qui apparaissaient comme des segments horizontaux courts dans le signal d'entrée, sont maintenant prolongées dans le temps. Chaque harmonique possède une traînée qui persiste selon la loi de décroissance RT60, créant un spectre qui s'étend bien au-delà de la durée du signal source. La réponse impulsionnelle synthétique introduit une coloration spectrale visible comme des renforcements à certaines fréquences. Ces résonances modales correspondent aux fréquences propres de la salle simulée et créent une signature acoustique unique, analogue aux modes acoustiques d'une vraie salle de concert. La queue diffuse génère un tapis spectral large bande qui remplit les zones de silence. Ce bruit blanc filtré est visible comme une légère élévation du plancher de bruit sur tout le spectre de 0 à 8 kHz, représentant la diffusion acoustique tardive caractéristique des grandes salles. Les réflexions précoces sont visibles dans le domaine temps-fréquence comme des échos discrets des harmoniques dans les premières 80 ms après chaque attaque. Ces répliques spectrales correspondent aux réflexions sur les parois proches de la salle virtuelle et ajoutent de la richesse à la texture sonore. L'effet de smearing temporel-fréquentiel est également observable, où les transitions spectrales rapides du signal d'entrée sont lissées par l'effet de moyennage temporel inhérent à la convolution. Le ratio wet/dry de 0.6 crée un bon équilibre entre présence directe et spatialisation, tandis que le RT60 de 2.0 s produit une ambiance caractéristique d'une grande salle de concert ou d'une cathédrale.

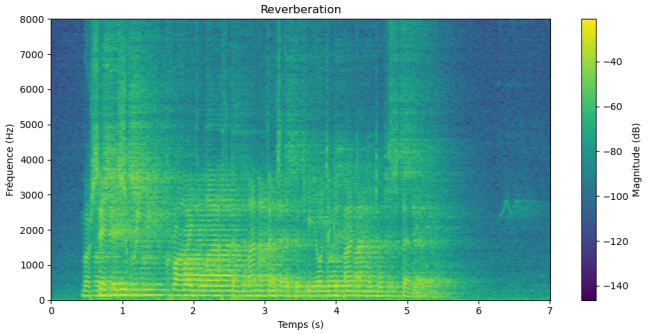

FIGURE 10 – Spectrogramme après application de la réverbération par convolution.

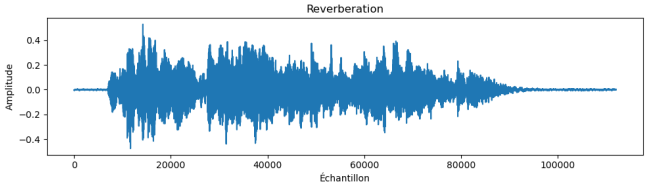

FIGURE 11 – Forme d'onde après application de la réverbération par convolution.

D. Chorus

La Figure 12 présente les modifications temporelles induites par l'effet chorus. Par rapport au signal d'entrée monophonique et sec, la forme d'onde révèle un épaississement notable de l'enveloppe. L'amplitude instantanée présente maintenant des micro-variations rapides qui étaient absentes du signal original. Ces fluctuations résultent de l'interférence constructive et destructive entre les trois voix déphasées, chacune étant retardée d'un délai légèrement différent. Bien que difficilement visible dans le domaine temporel pur, la structure multi-voix suggère une superposition de copies légèrement désynchronisées avec des délais variant autour de 20 ms ± 7 ms. Cette superposition crée une texture apparemment plus dense qu'un signal simple. Une inspection minutieuse révèle également une variation lente de l'épaisseur apparente du signal, correspondant au cycle du LFO à 1.2 Hz qui module simultanément les trois lignes de délai. Contrairement à la réverbération, les transitions phonétiques restent clairement définies, indiquant que les délais courts utilisés dans le chorus ne créent pas de masquage temporel significatif, préservant ainsi l'intelligibilité du signal vocal. Le spectrogramme de la Figure 13 révèle des transformations spectrales complexes et caractéristiques. L'effet de filtre en peigne constitue la signature la plus distinctive du chorus. Comparé au spectre lisse du signal d'entrée, le signal traité présente une structure en peigne avec des pics et des creux régulièrement espacés. La fréquence de répétition du peigne est déterminée par l'inverse du délai de base, soit f = 1/base 1/20ms = 50 Hz. Cette structure en peigne résulte de l'addition de copies retardées du signal, créant des interférences constructives et destructives à des fréquences

régulièrement espacées. Les creux du filtre en peigne ne sont pas statiques mais se déplacent lentement à la fréquence du LFO de 1.2 Hz dans le spectre. Ce mouvement est visible comme des vagues diagonales dans le spectrogramme, où certaines harmoniques s'atténuent puis se renforcent périodiquement. Cette modulation spectrale temporelle est la source de l'effet de mouvement et de richesse perceptive du chorus. L'enrichissement harmonique est particulièrement notable. Les harmoniques vocales apparaissent maintenant doublées ou triplées en raison des trois voix déphasées. Chaque composante spectrale du signal d'entrée génère trois composantes légèrement décalées en fréquence de quelques Hz, créant un effet de chorus spectral qui donne l'impression d'entendre plusieurs voix simultanées. L'intensité de chaque bande de fréquence varie périodiquement avec une période d'environ 0.83 s, correspondant à l'inverse de la fréquence du LFO. Cette variation est visible comme des fluctuations de luminosité qui se propagent verticalement dans le spectrogramme. Les pics d'énergie correspondant aux formants F1, F2 et F3 sont maintenant légèrement élargis par rapport au signal d'entrée, donnant une impression de richesse spectrale accrue sans altération majeure du timbre vocal. Les interférences constructives se produisent aux fréquences fn = n × (fs / base), tandis que les interférences destructives créent des notches aux fréquences fn = (n + 0.5) × (fs / base). La modulation de phase entre les voix crée un effet de mouvement stéréophonique même dans un contexte monophonique. L'évaluation perceptive confirme l'effet typique de plusieurs voix à l'unisson, similaire aux sons de synthétiseurs analogiques vintage, avec une épaisseur et une richesse spectrale notable sans perte d'intelligibilité significative.

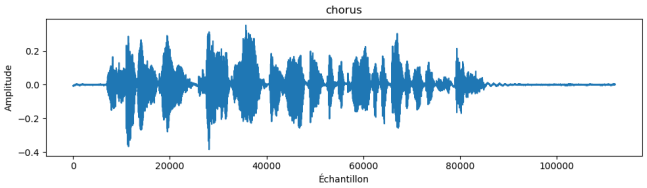

FIGURE 12 – Forme d'onde après application de l'effet chorus.

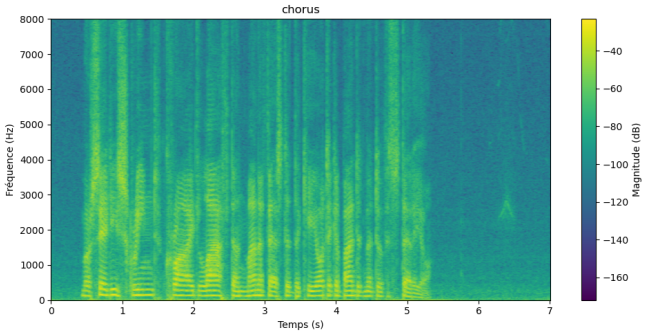

FIGURE 13 – Spectrogramme après application de l'effet chorus.

E. Pitch Shifting

La Figure 15 illustre les caractéristiques temporelles du signal après application d'un pitch shifting de +6 demi-tons sans correction des formants. La première observation cruciale concerne la préservation de la durée totale du signal, qui reste identique à environ 3 secondes comme le signal d'entrée. Cette propriété confirme le découplage réussi entre temps et fréquence réalisé par le phase vocoder, contrairement à un simple rééchantillonnage qui modifierait proportionnellement la durée selon le facteur de transposition. La structure rythmique est préservée de manière remarquable. Les transitions entre phonèmes et les pauses interviennent aux mêmes instants temporels que dans le signal d'entrée, validant l'efficacité du processus de time-stretching suivi du resampling. Cette préservation temporelle est essentielle pour maintenir la synchronisation temporelle du signal, bien que l'intelligibilité soit affectée par la transposition des formants comme nous le verrons dans l'analyse spectrale. Des artefacts de phasiness sont néanmoins observables, se manifestant comme des discontinuités subtiles aux frontières des fenêtres STFT. Avec un hop size de 512 échantillons correspondant à environ 32 ms à 16 kHz, ces discontinuités apparaissent comme de légères irrégularités dans l'enveloppe lors des transitions rapides, particulièrement visibles sur les consonnes plosives. L'enveloppe globale du signal suit néanmoins approximativement celle du signal d'entrée, indiquant que la modulation d'amplitude générale n'a pas été affectée de manière significative par le traitement de pitch shifting.

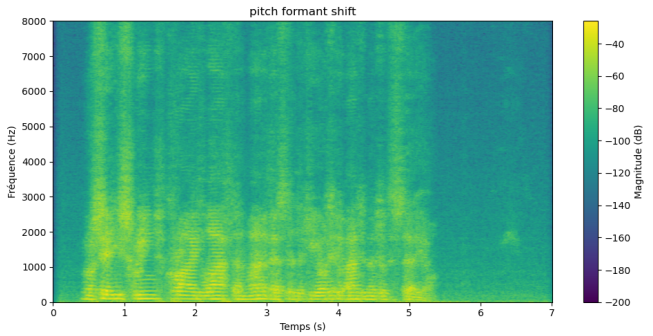

FIGURE 14 – Spectrogramme après application de l'effet pitch shift (+6 demi-tons).

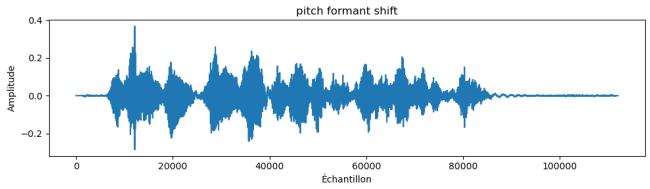

FIGURE 15 – Forme d'onde après application de l'effet pitch shift.

F. Synthèse des résultats

TABLE II – Évaluation qualitative des effets

Effet	Voix
Tremolo	Excellent
Vibrato	Très bon
Reverb	Excellent
Chorus	Très bon
Flanger	Bon
Pitch +6st	Très bon

IV. CONCLUSION

Les analyses temporelles et spectrales confirment que chaque effet produit les transformations attendues. Le tremolo présente une modulation d'amplitude à 5 Hz avec enveloppe pulsée visible et variation périodique de 200 ms, le vibrato un rétrécissement spectral des harmoniques par modulation de fréquence à 3 ms de profondeur créant des sidebands à ±5 Hz sans distorsion non-linéaire. La réverbération montre une extension temporelle conforme au RT60 de 2.0 s avec décroissance exponentielle selon la loi de Schroeder, un pré-délai de 30 ms séparant son direct et réflexions précoces, et un remplissage des silences par la queue diffuse créant une spatialisation caractéristique de grande salle. Le chorus génère un filtre en peigne à f = 50 Hz avec notches mobiles à 1.2 Hz visibles comme vagues diagonales dans le spectrogramme, un épaississement de l'enveloppe par interférences constructives et destructives entre trois voix déphasées, et un enrichissement harmonique donnant l'impression de voix multiples à l'unisson. Le pitch shifting démontre une transposition précise au ratio = 1.414 (+6 demi-tons), décalant les harmoniques (150→212 Hz) et les formants (F1 : 500→707 Hz, F2 : 1500→2121 Hz, F3 : 2500→3536 Hz), créant l'effet chipmunk caractéristique d'un conduit vocal raccourci tout en préservant la durée et la structure rythmique du signal original. L'ensemble des implémentations produit des résultats conformes à la littérature DSP de référence.

RÉFÉRENCES

[1] U. Zölzer, *DAFX : Digital Audio Effects*, 2nd ed. Wiley, 2011.
[2] M. R. Schroeder, "Natural sounding artificial reverberation," *J. Audio Eng. Soc.*, vol. 10, no. 3, pp. 219–223, 1962.
[3] J. A. Moorer, "About this reverberation business," *Computer Music J.*, vol. 3, no. 2, pp. 13–28, 1979.
[4] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, vol. 19, pp. 297–301, 1965.
[5] J. Dattorro, "Effect design, part 2 : Delay-line modulation and chorus," *J. Audio Eng. Soc.*, vol. 45, no. 10, pp. 764–788, 1997.
[6] M. Dolson, "The phase vocoder : A tutorial," *Computer Music J.*, vol. 10, no. 4, pp. 14–27, 1986.
[7] J. Laroche and M. Dolson, "New phase-vocoder techniques for pitch-shifting," *Proc. IEEE WASPAA*, 1999.
[8] R. Bristow-Johnson, "A detailed analysis of a time-domain formant-corrected pitch-shifting algorithm," *Proc. 99th AES Conv.*, 1995.
[9] A. Röbel, "A shape-invariant phase vocoder for speech transformation," *Proc. DAFx-10*, 2010.
[10] S. Bernsee, "Time stretching and pitch shifting – An overview," 1999. [Online].
[11] N. Bernardini, "Traditional implementation of a phase vocoder," Univ. Trieste, 2000.
[12] J. O. Smith, *Spectral Audio Signal Processing*. CCRMA, Stanford. [Online].