

UE SI340 - Traitement des Signaux Audiofréquences -
Effets et Réverbération artificielle

Gaël RICHARD

23 mai 2007

Table des matières

| | |
|---|----------|
| V Effets et réverbération artificielle¹ | 1 |
| V.1 Introduction | 1 |
| V.2 Compression/Expansion. Limiteur | 1 |
| V.2.1 Estimation du niveau du signal en entrée | 2 |
| V.2.2 Actualisation des paramètres de contrôle dynamique | 3 |
| V.2.3 Amplificateur à gain contrôlé | 4 |
| V.2.4 Implémentation et utilisation | 6 |
| V.3 Flanging/Phasing | 7 |
| V.3.1 Définition et origine | 7 |
| V.3.2 Le "Flanging" | 7 |
| V.3.3 Le "Phasing" | 11 |
| V.4 Modifications de l'échelle temporelle et fréquentielle, harmoniseur | 11 |
| V.4.1 Définitions | 11 |
| V.4.2 La technique de la mémoire circulaire | 13 |
| V.4.3 Harmoniseur | 15 |
| V.5 Le "chorusing" | 16 |
| V.5.1 Implémentation par changement de fréquence | 16 |
| V.5.2 Implémentation par filtre passe-tout à retard variable | 16 |
| V.6 Distorsion | 19 |
| V.6.1 Idée générale | 19 |
| V.6.2 Implémentation à l'aide de polynômes de Chebyshev | 20 |
| V.7 Réverbération: introduction | 20 |
| V.7.1 Canal acoustique | 20 |
| V.7.2 Réverbération en lieu clos | 20 |
| V.7.3 Auralisation | 22 |
| V.7.4 Modèles physiques et perceptifs en acoustique des salles | 22 |
| V.8 Algorithmes de réverbération artificielle | 23 |
| V.8.1 Filtres en peigne et filtres passe-tout | 24 |
| V.8.2 Le réverbérateur de Schroeder | 25 |
| V.8.3 Simulation d'un temps de réverbération dépendant de la fréquence | 27 |
| V.8.4 Systèmes unitaires | 27 |
| V.8.5 Filtres réverbérants à bouclage unitaire | 28 |
| V.8.6 Contrôle des réflexions précoces | 30 |

1. Chapitre principalement écrit par J.M.Jot et J. Laroche

Chapitre V

Effets et réverbération artificielle¹

V.1 Introduction

Ce chapitre regroupe deux domaines important en traitement du signal audio. Dans une première partie, nous décrivons un certain nombre de transformations du signal audio. Certaines de ces transformations sont principalement réalisées dans un but d'efficacité d'un système de radio-diffusion ou d'un système de restitution sonore (ce sera le cas des systèmes de compression/expansion) alors que d'autres sont réalisées dans le but de changer les caractéristiques du son. Ces transformations permettent par exemple d'ajouter des "effets" artificiels au son original (ce sera le cas du *flanger*, du *phasing* ou encore de la *distorsion*). La *réverbération artificielle* entre également dans ce cadre même si on ne cherche pas ici à modifier le signal dans le but d'obtenir un son au timbre original mais plutôt à simuler la façon dont une source sonore serait perçue dans un espace naturel réverbérant (comme par exemple une cathédrale ou une salle de concert).

Ce chapitre est ainsi organisé en deux parties principales:

- La première partie est consacrée aux effets audio. Ce chapitre n'inclura pas la description des effets de compression/dilatation temporelle et fréquentielle à partir du vocodeur de phase ou des méthodes plus spécifiques au traitement de la parole qui sont décrites dans un autre chapitre.
- La seconde partie est dédiée aux algorithmes de réverbération artificielle

V.2 Compression/Expansion. Limiteur

Les dispositifs de compression/expansion et de limitation de dynamique sont utilisés depuis très longtemps en analogique pour contrôler la dynamique des signaux (rapport entre la puissance du signal le plus fort et celle du signal le plus faible). En effet, si la dynamique d'une salle de concert dépasse parfois $100dB$, celle d'une émission radio en modulation de fréquence ne peut dépasser $70dB$. Pour que les pianissimi ne se retrouvent pas noyés dans le bruit de fond, il est nécessaire de diminuer la dynamique du signal pour la faire loger dans celle du canal de transmission. Cette opération est réalisée avant la diffusion par un "compresseur de dynamique". A l'inverse, après réception, il est nécessaire de restaurer la dynamique originale, opération réalisée par une "expansion de dynamique".

Un limiteur de dynamique est un dispositif qui empêche un signal de dépasser une puissance prédéfinie. Il introduit à ce titre une distorsion qui est cependant préférable à la distorsion qui

1. Chapitre principalement écrit par J.M.Jot et J. Laroche

résulterait d'un écrêtage ou d'une saturation (voir chapitre V.6).

Les trois types de contrôle de dynamique (compression/expansion et limitation) fonctionnent selon le même principe, mais différent par leur réglage. Les trois principales étapes de leur implémentation sont:

1. Calcul du niveau du signal en entrée.
2. Actualisation des paramètres de contrôle dynamique.
3. Contrôle dynamique du niveau du signal en échelle logarithmique.

Le niveau calculé lors de la première étape sert à contrôler un amplificateur à gain variable appliqué sur le signal d'entrée, selon le schéma donné figure V.1.

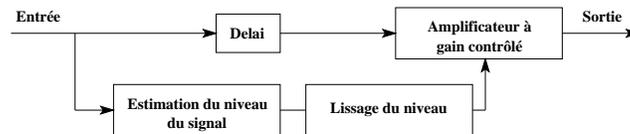


FIG. V.1 – Schéma du contrôle dynamique du niveau du signal

On pourra se référer à [16] pour tous les détails.

V.2.1 Estimation du niveau du signal en entrée

On retient deux mesures du niveau du signal en entrée:

Valeur efficace C'est l'amplitude du signal continu de même puissance. On la calcule par:

$$P_{eff} = \sqrt{\frac{1}{T} \int_{t-T}^t x^2(u) du} \quad \text{ou en discret:} \quad P_{eff} = \sqrt{\frac{1}{N} \sum_{i=n-N+1}^n x_i^2}$$

L'entier N définit l'horizon de la mesure. Cette mesure est également appelée *RMS* (root-mean-square). Elle présente l'avantage de ne pas dépendre des phases relatives des composantes du signal. C'est la mesure qui est retenue pour la compression/expansion de dynamique. Pour une sinusoïde d'amplitude A , la valeur efficace calculée sur une période vaut $A/\sqrt{2}$.

Niveau crête à crête C'est la mesure de l'amplitude maximale du signal. On la calcule par:

$$P_{crete} = \max_{n-M < i < n} |x_i|$$

C'est la mesure utilisée par les limiteurs (puisque leur but est d'éviter l'écrêtage du signal).

On peut définir pour un signal le *facteur de crête* qui est le rapport du niveau crête à crête à la valeur efficace. C'est un nombre toujours plus grand que 1 qui reflète la présence de pics dans le signal. Pour une sinusoïde, le facteur de crête est $\sqrt{2}$.

Le niveau d'entrée est donc estimé en valeur efficace pour la compression/expansion, et en valeur crête à crête pour les limiteurs. Cette estimation va être maintenant utilisée pour contrôler l'amplification du signal d'entrée.

V.2.2 Actualisation des paramètres de contrôle dynamique

Le niveau d'entrée est destiné à contrôler l'amplificateur de sortie. Il est cependant nécessaire de le lisser pour éviter les variations trop brusques de gain d'amplification, en particulier lors des montées et des chutes brutales de niveau. Ce comportement dynamique de type "oubli exponentiel" est généralement contrôlé par deux paramètres, le temps de montée et le temps de relaxation. On utilise généralement une implémentation de type filtre passe-bas récursif:

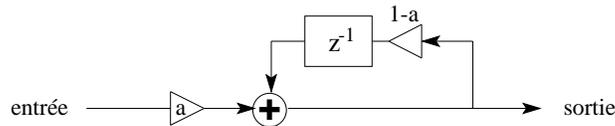


FIG. V.2 – Implémentation d'un système de contrôle dynamique sous forme d'un filtre passe-bas récursif

Sa fonction de transfert

$$H(z) = \frac{a}{1 - (1 - a)z^{-1}}$$

permet de réaliser le filtrage passe-bas nécessaire.

Temps de montée Lorsque le niveau du signal présente une "marche d'escalier", le système de contrôle transforme la montée verticale en une courbe de type exponentiel. Le temps mis par le signal de gain pour atteindre le niveau réel à moins de $2dB$ près est appelé temps de montée.

Temps de relaxation Lorsque le niveau du signal chute verticalement, le système de contrôle transforme la chute verticale en une courbe de type exponentiel. Le temps mis par le signal de gain pour atteindre le niveau réel à moins de $2dB$ près est appelé temps de relaxation. Il est en général plus long que le temps de montée.

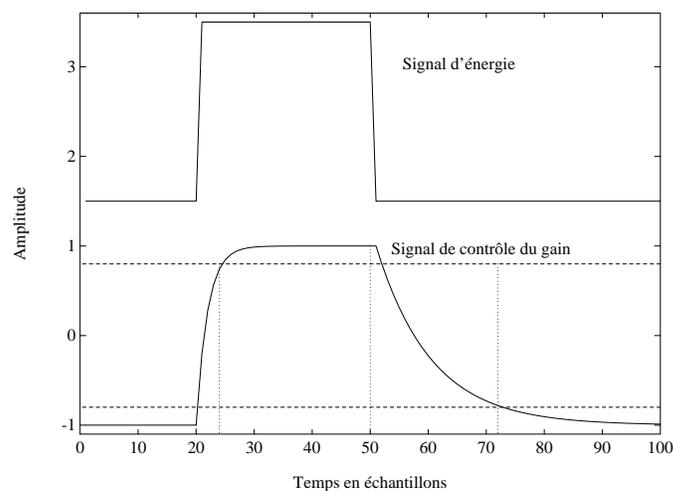


FIG. V.3 – Illustration du temps de montée et du temps de relaxation du gain d'amplification

Les deux paramètres temps de montée et temps de relaxation définissent les caractéristiques dynamiques du système. Le temps de montée est de l'ordre de quelques millisecondes (pour réagir rapidement aux attaques) et le temps de relaxation de l'ordre de quelques centaines de millisecondes (pour qu'il y ait effectivement un lissage). La réponse du filtre H ci-dessus à un signal en marche d'escalier permet d'adapter le paramètre a pour obtenir le temps de montée (ou de relaxation) voulu.

V.2.3 Amplificateur à gain contrôlé

Le signal d'entrée est amplifié par un amplificateur dont le gain est contrôlé par l'estimation lissée du niveau du signal d'entrée. La courbe donnant le gain d'amplification en dB en fonction du niveau estimé du signal définit les caractéristiques statiques du système. C'est par ces caractéristiques que les trois systèmes de compression, d'expansion et de limitation diffèrent.

Compression de dynamique Lors de la compression de dynamique, les signaux de fort niveau sont moins amplifiés que les signaux de faible niveau. La figure V.4 présente la courbe donnant la valeur du gain d'amplification en fonction du niveau du signal d'entrée en dB ainsi que le niveau du signal en sortie en fonction du niveau en entrée pour un compresseur de dynamique. Les échelles sont en décibels pour garantir une "linéarité" subjective.

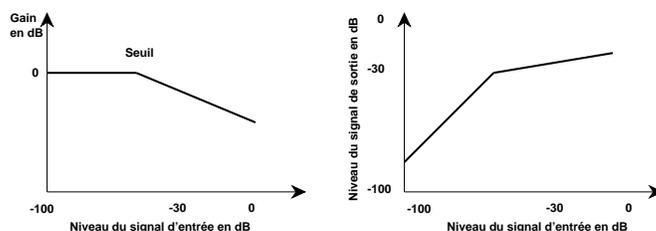


FIG. V.4 – Valeur du gain d'amplification en fonction du niveau du signal d'entrée en dB (à gauche); Niveau du signal en sortie en fonction du niveau en entrée pour un compresseur de dynamique (à droite)

La courbe de compression est caractérisée par un seuil et un taux de compression.

- Le seuil est la valeur du niveau d'entrée en dB à partir de laquelle le gain n'est plus constant (le système commence à compresser).
- Le taux de compression est par définition la pente de la caractéristique entrée/sortie dans la région de compression. C'est un nombre sans dimension.

Si l'on désigne par X_{eff} et Y_{eff} les valeurs efficaces *en dB* du signal en entrée et en sortie, et si l'on désigne par $g(X_{eff})$ le gain d'amplification en dB (qui dépend du niveau du signal en entrée), on a

$$Y_{eff} = X_{eff} + g(X_{eff}) \quad \text{d'où} \quad \frac{dY_{eff}}{dX_{eff}} = 1 + \frac{dg(X_{eff})}{dX_{eff}} = 1 + \alpha$$

où α représente la pente du gain (figure de gauche). On voit que si le gain est linéaire en X_{eff} de pente α , la caractéristique entrée/sortie est aussi linéaire avec une pente égale à $1 + \alpha$.

Pour les systèmes courants de compression de dynamique, le rapport de compression est de l'ordre de 1/2. On constate sur la figure ci-dessus que la dynamique du signal en sortie est inférieure à celle du signal en entrée, ce qui permet par exemple de l'adapter à un canal de transmission bruité.

Pour qu'il soit possible de retrouver le signal original, il faut que la caractéristique d'entrée-sortie soit *strictement monotone*, ce qui implique que la pente du gain soit inférieure à 1 en valeur absolue (voir équation ci-dessus). En pratique, l'utilisation de taux de compression trop importants entraîne lors de l'expansion d'importantes distorsions.

Expansion de dynamique C'est le système inverse du système précédent. Ses caractéristiques sont les suivantes (voir figure V.5).

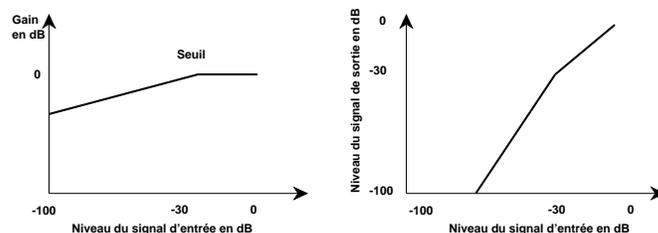


FIG. V.5 – Valeur du gain d'amplification en fonction du niveau du signal d'entrée en dB (à gauche); Niveau du signal en sortie en fonction du niveau en entrée (à droite) pour une expansion de dynamique

De la même façon, on définit un seuil d'expansion, (niveau du signal d'entrée en dB au delà duquel le signal est amplifié avec un gain constant), et un taux d'expansion, pente de la caractéristique entrée-sortie dans la région d'expansion. On voit que la caractéristique entrée-sortie présente une pente supérieure à 1: lorsque le taux d'expansion est trop important, une petite erreur d'estimation du niveau du signal d'entrée se traduit par une grande variation du niveau du signal de sortie: l'utilisation de taux d'expansion trop élevés conduit à d'importantes distorsions.

Limiteur Ce système est utilisé lorsqu'on veut éviter que le niveau du signal dépasse une valeur prédéfinie. Sa caractéristique est donnée sur la figure V.6:

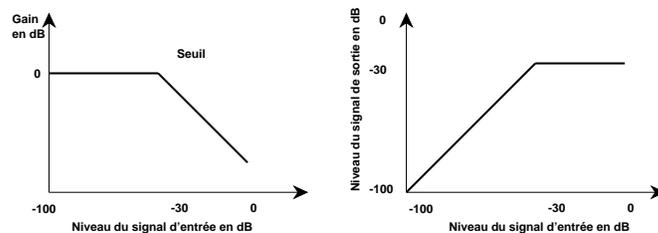


FIG. V.6 – Valeur du gain d'amplification en fonction du niveau du signal d'entrée en dB (à gauche); Niveau du signal en sortie en fonction du niveau en entrée (à droite) pour un limiteur

Le gain à partir du seuil de limitation a une pente de valeur -1 qui donne au signal de sortie un niveau constant quel que soit celui du signal d'entrée. Bien évidemment, cette opération n'est pas réversible puisque la caractéristique entrée/sortie n'est pas strictement monotone.

V.2.4 Implémentation et utilisation

L'implémentation en temps réel des dispositifs de contrôle de dynamique peut se faire relativement aisément. Les conversions linéaire/logarithmique logarithmique/linéaire nécessaires au calcul des gains en dB réclament cependant un soin particulier [16].

La figure (V.7) donne un exemple de modification dynamique d'un signal carré. Sur le graphique du haut de la figure, le signal est correctement traité, on constate que le temps de montée est suffisamment court pour que le front montant du signal original soit instantanément 'tassé'. On constate sur le front descendant le temps de relaxation plus long qui se traduit par un gain croissant lentement.

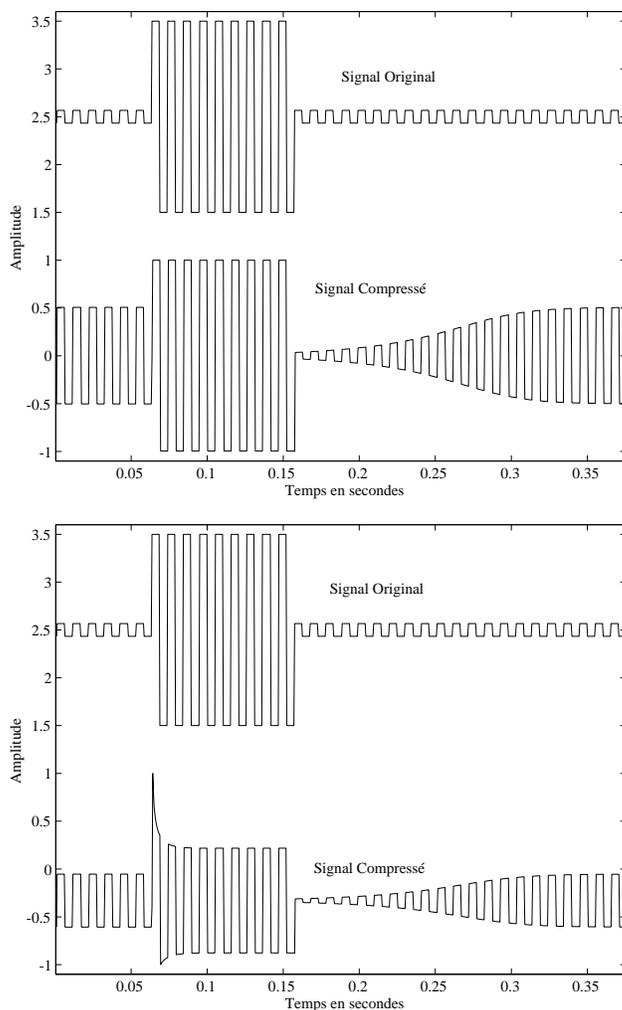


FIG. V.7 – Exemples de compression de la dynamique d'un signal. A gauche, compression d'un signal de taux 0.8. A droite, même opération mais avec un temps de montée trop long (l'échelle des amplitudes est différente)

Sur le graphique du bas, le temps de montée est trop long, le gain d'amplification est ajusté trop lentement: pendant un temps très bref, le signal possède une forte amplitude.

Les dispositifs de compression/expansion ont été très utilisés en analogique car ils permettaient de diminuer le bruit de fond des supports magnétiques. Ainsi les Dolby B et C ainsi que le

DBX sont basés sur la compression réversible de la dynamique du signal (avec préaccentuation des aiguës à l'enregistrement et désaccentuation à l'écoute).

Avec l'avènement du numérique, ces dispositifs sont surtout utilisés pour assurer la liaison entre le domaine numérique et le domaine analogique: transmission d'un signal à forte dynamique au moyen d'un canal analogique bruité (radio), enregistrement magnétique traditionnel (Cassette) d'un signal numérique à grande dynamique, etc...

Les compresseurs de dynamique à fort taux de compression sont couramment utilisés par les guitaristes pour obtenir un effet de "sustain" qui permet à leur guitare de sonner indéfiniment ou presque: lors de l'attaque, la compression fait disparaître le transitoire de forte énergie, puis lors de la résonance, le gain d'amplification croît à mesure que le son décroît, produisant en sortie un signal d'amplitude pratiquement constante.

Enfin les limiteurs sont toujours utilisés comme dispositifs de protection, ou afin d'éviter un écrêtage par saturation: Les limiteurs sont très utilisés en radiodiffusion où ils servent à éviter la saturation des appareils d'émission. Le signal délivré à l'émetteur ne doit en général pas dépasser un niveau maximal de modulation (sous peine d'écrêtage), et le limiteur permet de s'assurer que ce niveau maximal n'est jamais dépassé. Les compresseurs sont également largement utilisés par certaines radios (que nous ne nommerons pas...) pour 'tasser' la dynamique du signal, ce qui permet d'émettre à un niveau constamment élevé.

V.3 Flanging/Phasing

V.3.1 Définition et origine

Le terme flanging vient de l'origine analogique de cet effet: les disc jockeys avaient découvert que s'ils jouaient le même disque simultanément sur deux platines en mélangeant les deux sorties, ils pouvaient obtenir un effet spécial à ce jour inconnu. Il suffisait de placer le pouce sur le bord de chaque plateau ("flange" en anglais), puis de ralentir tour à tour les deux platines qui se retrouvaient tantôt synchronisées, tantôt légèrement déphasées. Lorsque les deux sorties étaient également amplifiées, le son résultant prenait une coloration spéciale qui variait au cours du temps. C'est l'origine du terme "flanging".

Si l'on modélise ce qui se passe lors du "flanging" analogique, on s'aperçoit que les signaux arrivant des deux platines sont retardés d'un délai qui varie au cours du temps. Lorsque ces signaux sont ajoutés, certaines composantes sinusoïdales sont annulées car elles se retrouvent en opposition de phase. Le "flanging" introduit donc des trous dans le spectre du signal original, et leur position varie au cours du temps. On pourra donc définir:

Flanging: l'effet par lequel on introduit des zéros équirépartis en fréquence dans le spectre original.

Phasing: le même effet lorsque les zéros ne sont plus équirépartis.

V.3.2 Le "Flanging"

Implémentation

L'opération de *flanging* peut être réalisée en reproduisant en numérique le système analogique décrit ci-dessus [29, 2]. On implémente donc un retard de longueur variable qui est ajouté au signal original, conformément au schéma représenté sur la figure V.8:

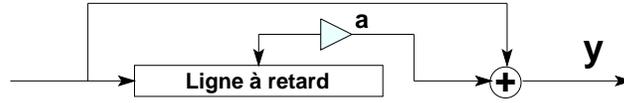


FIG. V.8 – Schéma d'implémentation du flanging

où le retard subit une variation périodique dans le temps. Si l'on suppose le retard fixé à une valeur donnée p (en nombre d'échantillons), la fonction de transfert du système est simplement:

$$H(z) = 1 + a z^{-p} \quad \text{d'où} \quad |H(e^{j\omega})| = \sqrt{1 + a^2 + 2a \cos(p\omega)}$$

Le paramètre a règle le niveau du signal retardé. La fonction de transfert du système figé présente p minima aux multiples impairs de $Fe/2p$:

En théorie, lorsque le paramètre de retard p varie avec le temps, on ne peut plus parler de fonction de transfert, puisque le filtre n'est plus invariant dans le temps. Toutefois si cette variation est lente, on constate que le système introduit des régions à faible énergie autour de fréquences en rapport harmonique dont la localisation varie au cours du temps.

En général, le retard varie en fonction du temps selon une courbe en dent de scie, avec une période longue (de l'ordre de la seconde). Le résultat est une coloration du signal périodique de même période.

Le paramètre a permet de régler la profondeur de l'effet: lorsque a vaut zéro, l'effet disparaît, lorsque a vaut un, l'effet est maximal. Le rapport en dB du maximum au minimum du module de la fonction de transfert peut s'exprimer en fonction du paramètre a :

$$R_{dB} = 20 (\log_{10}(1 + a) - \log_{10}(1 - a))$$

où l'on s'aperçoit que pour obtenir un contrôle "linéaire en dB" de l'effet dans la région intéressante (a voisin de 1), il est utile de le paramétrer par $\alpha = \log_{10}(1 - a)$ soit $a = 1 - 10^{-\alpha}$

Le problème des retards non-entiers

Le retard p prend toujours une valeur entière, ce qui conduit à des effets désagréables. Si par exemple le retard passe de la valeur 3 à la valeur 4, la fonction de transfert du filtre change radicalement, ce qui peut produire des discontinuités dans le signal de sortie. En effet, si le signal en entrée est une sinusoïde de pulsation $3\pi/4$, lorsque p vaut 3 et $a = 1$, son amplitude en sortie est multipliée par 1.84. Lorsque p vaut 4, son amplitude en sortie devient nulle.

Pour rendre le comportement du flanger plus régulier, il faut utiliser des "retards non-entiers". Pour cela, il faudrait pouvoir disposer de la valeur du signal *entre* les échantillons. Le problème qui se pose ressemble un peu à celui du changement de fréquence d'échantillonnage mais se révèle plus complexe à résoudre.

En théorie, pour déterminer la valeur du signal entre les échantillons, il faut faire la convolution du signal discret par un sinus cardinal (théorème de reconstruction). En pratique, c'est une opération beaucoup trop chère pour les applications temps-réel: non seulement la convolution est à support infini, mais les coefficients du sinus cardinal doivent être recalculés pour chaque nouvelle valeur du retard p .

En pratique, on utilise une interpolation linéaire. La valeur intermédiaire du signal est calculée par interpolation linéaire entre les deux échantillons qui l'entourent. Pour une valeur de τ fixée,

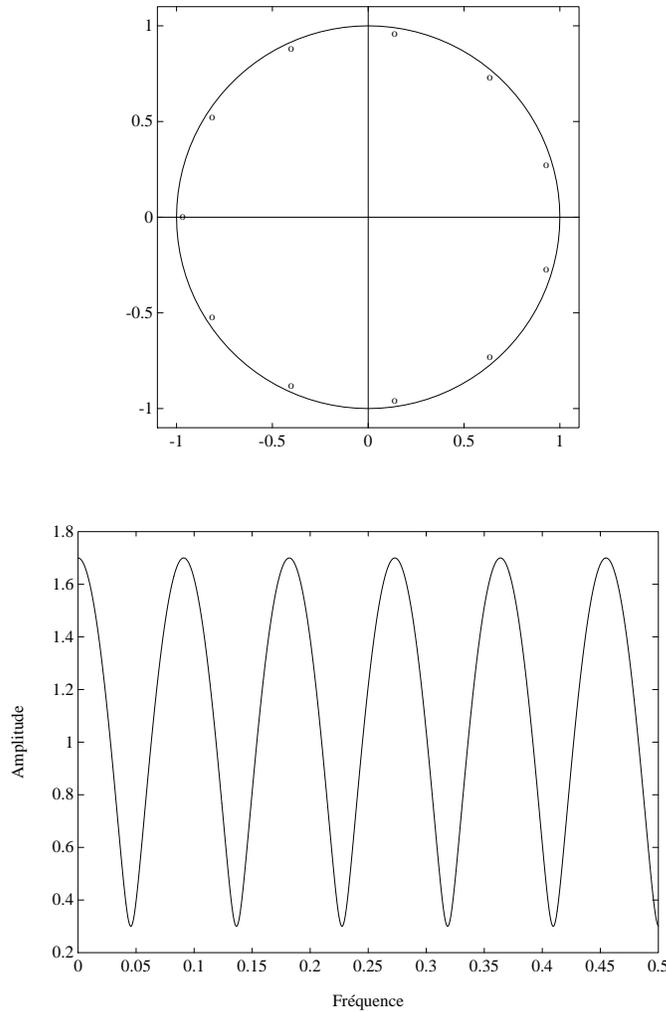


FIG. V.9 – *Fonction de transfert du système figé*

il s'agit donc d'un filtrage linéaire de fonction de transfert:

$$G(z) = (1 - \tau) + \tau z^{-1}$$

où τ représente la valeur du retard (en fraction d'échantillon). Le filtre en peigne avec retard non-entier a donc pour fonction de transfert (pour un retard fixé)

$$H(z) = 1 + a[(1 - \tau)z^{-p} + \tau z^{-p-1}]$$

Le retard non entier diminue la profondeur du peigne dans les hautes fréquences ainsi qu'on peut le constater sur la figure V.10 pour un retard $p = 6$, $a = 0.9$ et $\tau = 0.2$ puis $\tau = 0.8$ (c'est à dire un retard non entier de 6.2 puis 6.8 échantillons).

L'interpolation linéaire permet d'éviter les problèmes de discontinuité dûs aux variations brusques du retard p , puisque celui-ci peut maintenant varier "continûment". En contrepartie, interpoler linéairement diminue l'effet de phasing en haute fréquence. Il faut noter cependant que l'interpolation linéaire, malgré son faible coût en calculs possède quelques propriétés intéressantes pour le signal audio. Voir [19] pour une discussion de l'utilisation de l'interpolation linéaire dans le cadre de la synthèse.

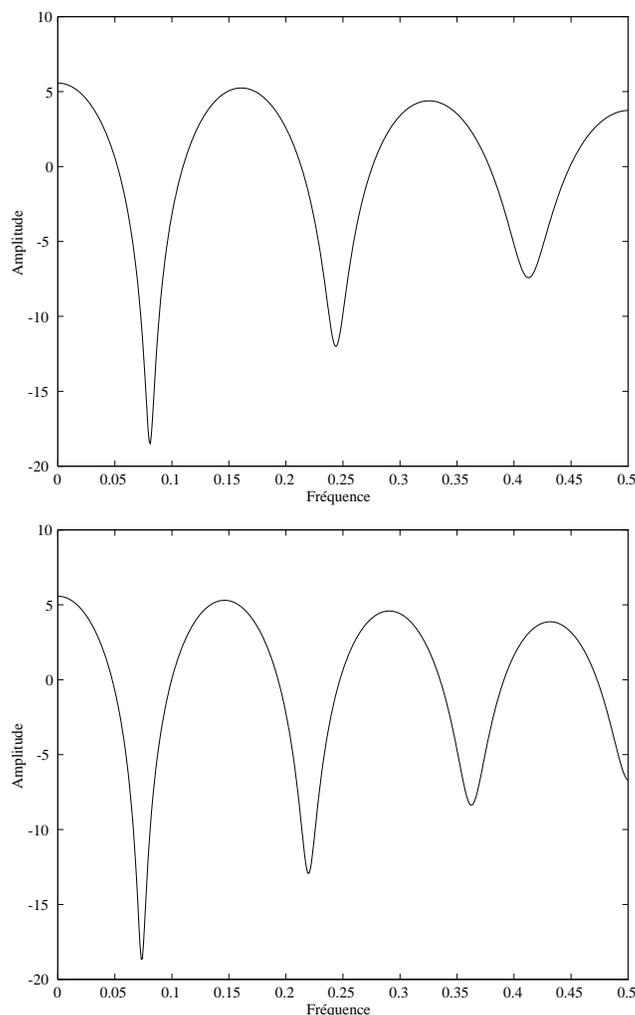


FIG. V.10 – *Le retard non entier diminue la profondeur du peigne dans les hautes fréquences (un retard non entier de 6.2 à gauche puis 6.8 échantillons, à droite)*

Implémentation en passe-tout

Nous avons vu que par définition, l'effet de "phasing" introduit des zones de faible énergie dans le spectre, à des fréquences qui ne sont plus équiréparties. A l'origine, on cherche à éviter l'équirépartition des fréquences pour éviter l'annulation pure et simple de certains signaux harmoniques dont le fondamental est voisin de $Fe/2p$ [29]. Une façon de réaliser cela consiste à remplacer la ligne à retard par un ensemble de filtre passe-tout en série, selon le schéma donné figure V.11:

On peut montrer que si le filtre passe-tout constitué des cellules en série possède p pôles dans le cercle unité, le filtre de phasing correspondant au schéma ci-dessus avec $a = 1$ possède p zéros *sur* le cercle unité. On peut le voir facilement pour une cellule passe-tout du second ordre réelle: la phase en fonction de ω est strictement décroissante, elle vaut zéro pour $\omega = 0$ et -2π pour $\omega = \pi$. Puisqu'elle est continue en ω , il existe une valeur ω_0 pour laquelle la phase de la cellule passe-tout vaut $-\pi$. Une sinusoïde de fréquence ω_0 se retrouve en sortie du passe-tout avec la même amplitude, et un déphasage de π . Lorsqu'elle est ajoutée à elle même, l'opposition

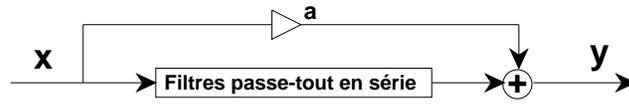


FIG. V.11 – Implémentation passe-tout

de phase génère un signal nul.

Ainsi, la série de cellules passe-tout possède une réponse en phase du type de celle donnée figure V.12 (en haut) où les pointillés sont les multiples impairs de $-\pi$. La figure de droite de la figure V.12 présente la réponse en amplitude du flanger pour cette configuration du filtre passe-tout. On constate que les fréquences auxquelles la réponse en amplitude s'annule ne sont plus multiples d'un fondamental, comme dans le cas du flanger.

L'implémentation en cellules du second ordre permet de contrôler dynamiquement chaque "notch" (entaille) dans la réponse en amplitude du flanger. De plus, cela permet aussi d'éviter le problème de l'interpolation rencontré ci-dessus: pour obtenir un filtrage variant dans le temps, on n'utilise plus un retard variable (qui nécessite une interpolation), mais on modifie les coefficients des cellules passe-tout. On pourrait également réaliser le même type d'effet à l'aide de simple filtre FIR (en positionnant p zéros sur le cercle unité), mais à cause de la troncature des coefficients des filtres, on n'est plus certain que les zéros sont exactement sur le cercle unité.

V.3.3 Le "Phasing"

Comme l'effet de flanging, le phasing est basé sur l'utilisation d'un filtre en peigne variant au cours du temps. Par opposition avec le flanging, ce n'est plus la durée du retard qui varie périodiquement mais le gain a de la boucle. Ainsi, selon que a est proche de zéro, ou voisin de 1, la courbe de réponse du "phaser" est plate ou possède des minima équidistants dont les positions sont constantes au cours du temps, mais dont la profondeur varie avec le paramètre a . Contrairement au flanging, l'effet de phasing ne nécessite pas d'interpolation.

V.4 Modifications de l'échelle temporelle et fréquentielle, harmoniseur

V.4.1 Définitions

On s'intéresse dans cette partie aux méthodes permettant de réaliser indépendamment une modification de l'échelle temporelle ou fréquentielle d'un signal:

- La modification de l'échelle temporelle permet d'altérer arbitrairement la durée d'un signal sans en modifier (si possible) le contenu fréquentiel. Bien sûr, la lecture du signal à une fréquence d'échantillonnage différente de celle de l'enregistrement conduit bien à une modification de la durée du signal, mais altère également son contenu fréquentiel (ex: disque 45 tours joué en 33 tours), ce que l'on cherche à éviter.
- La modification de l'échelle fréquentielle ("pitch shifting") est l'opération duale de la précédente, et consiste à modifier la hauteur d'un son donné, *sans en modifier la durée*. On peut distinguer deux types de transposition:
 - 1) Le changement de hauteur *sans modification des formants*. Dans le cas de la voix, ils

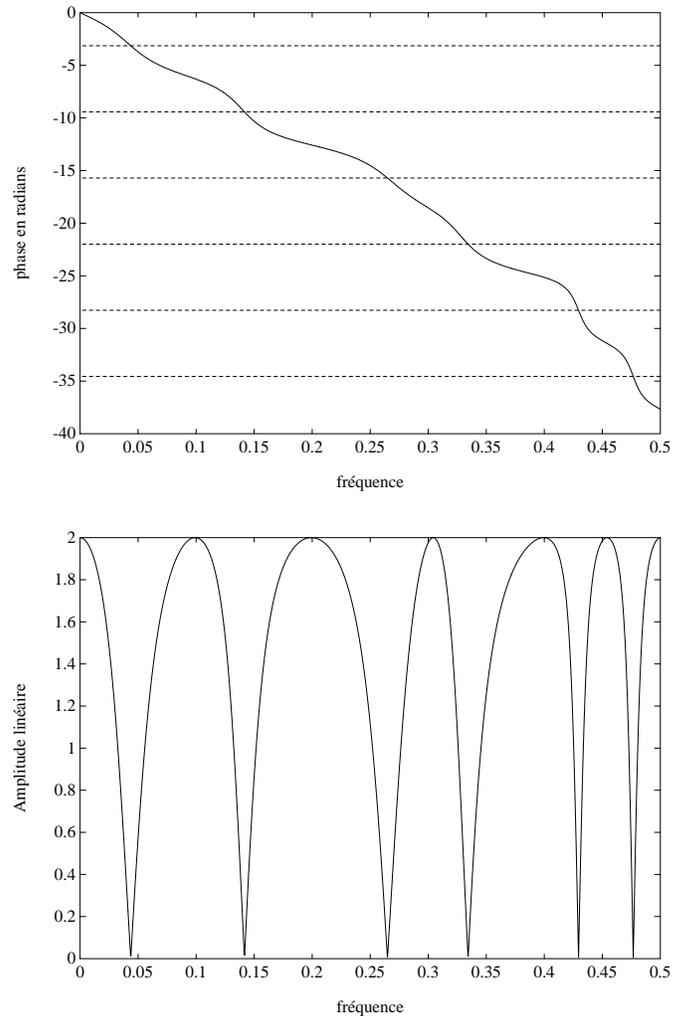


FIG. V.12 – Réponse en phase et en amplitude du flanger pour une configuration utilisant un filtre passe-tout

est souvent utile d'avoir un contrôle indépendant du fondamental et des formants. Les méthodes appartenant à cette catégorie sont capable de modifier le fondamental en respectant les positions et les caractéristiques des formants.

2) Le "changement de hauteur simple" ne respecte pas les positions des formants, mais effectue une homothétie de l'axe des fréquences. Ce sont en général des méthodes moins spécifiques de la parole que les précédentes.

Les applications de ces méthodes sont multiples:

- Les techniques de modification de l'échelle temporelle sont utilisées dans le domaine du traitement de la parole (synthèse à partir du texte) pour le rendu de la prosodie. Elles sont également très utiles pour les problèmes de post-synchronisation vidéo/son (par exemple pour ajuster la durée d'une bande son à celle d'un reportage vidéo), et pour la post-synchronisation des dialogues avec l'image (industrie du cinéma). Elles trouvent également des applications dans le domaine de l'enseignement des langues (possibilité de modifier la

rapidité d'élocution), ou du contrôle de bande audio (possibilité d'accélérer un enregistrement en maintenant l'intelligibilité de la parole).

- Les techniques de modification de l'échelle fréquentielle sont utilisées principalement pour la synthèse de la parole à partir du texte (rendu de la prosodie).

Les méthodes permettant de réaliser une modification de l'échelle temporelle ou fréquentielle se répartissent en deux catégories: méthodes paramétriques, reposant sur un modèle de signal précis (par exemple, modèle sinusoïdal), et méthodes non-paramétriques (où il n'est fait aucune hypothèse sur la nature du signal traité). Ces méthodes sont décrites pour partie dans le module PAROLE (méthode TD-PSOLA notamment) et pour partie dans le module MSA (méthode à base de vocodeur de phase) de la brique PAMU. Ainsi, nous ne parlerons ici que de méthodes particulières telle la technique de la mémoire circulaire décrite ci-dessous.

V.4.2 La technique de la mémoire circulaire

La technique de la mémoire circulaire est la plus simple et la plus ancienne des techniques de modification de l'échelle temporelle et fréquentielle [2]. Il s'agit également d'une méthode fonctionnant dans le domaine temporel.

L'origine analogique

Cette technique dérive d'un système analogique proposée dans les années 50 [7]. Elle consiste à utiliser un magnétophone muni d'une tête rotative. La bande en boucle fermée s'enroule sur la moitié du cylindre (comme pour les magnétoscopes et les DAT) et défile à vitesse constante. Le cylindre est muni de deux têtes de lecture diamétralement opposées dont les signaux sont mélangés avec un gain identique. Il est possible de contrôler le sens de rotation et la vitesse du cylindre.

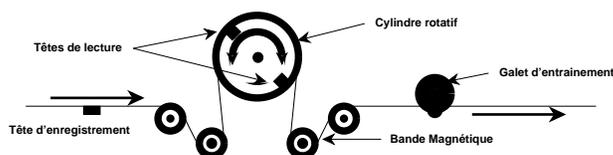


FIG. V.13 – Système de lecture/écriture pour l'implémentation de la technique de la mémoire circulaire

Lorsque le cylindre est immobile, la bande défile de façon identique devant la tête d'enregistrement et devant l'une des têtes de lecture. Le signal lu est donc identique au signal enregistré (aux erreurs d'enregistrement près).

Lorsque le cylindre tourne en sens inverse du défilement de la bande, la vitesse relative V_r de défilement de la bande par rapport à la tête de lecture est supérieure à sa vitesse de défilement absolue V_a . Pendant la durée du contact entre la tête de lecture et la bande, le signal est donc lu plus rapidement qu'il n'a été enregistré, ce qui correspond à une dilatation de l'axe des fréquences. La présence de deux têtes assure la continuité grâce à un "cross-fade" naturel (lorsqu'une tête quitte la bande, l'autre s'en rapproche, de sorte que le signal total ne diminue pas). On remarque que certaines portions du signal peuvent être lues *deux ou plusieurs fois*, en fonction de la vitesse de rotation de la tête. C'est cette relecture qui permet de conserver la durée du signal.

A l'inverse, lorsque le cylindre tourne dans le sens de défilement de la bande, le contenu en

fréquence du signal est contracté vers l'origine puisque la bande est lue à une vitesse moindre qu'elle n'est enregistrée. Dans ce cas, certaines portions du signal peuvent ne pas être lues du tout.

Le rapport de l'homothétie en fréquence s'exprime par:

$$\alpha = \frac{V_r}{V_a} = \frac{V_a + R \Omega_{cylindre}}{V_a}$$

où V_a est la vitesse de défilement de la bande devant la tête d'enregistrement, V_r la vitesse relative de la bande par rapport à la tête de lecture, $\Omega_{cylindre}$ la vitesse de rotation du cylindre en radians s^{-1} , et R le rayon du cylindre.

Dans tous les cas, l'alternance régulière des deux têtes se traduit par un "bruit" périodique de fréquence $\Omega_{cylindre}/\pi$.

Les modifications de l'échelle temporelle du signal sont obtenues par exemple en enregistrant le signal une première fois sur la bande, puis en le rejouant avec une vitesse de défilement de bande multipliée par un facteur α . En l'absence de rotation de la tête de lecture, la hauteur du signal est bien sûr multipliée par le facteur α , ce que l'on cherche à éviter. On compense donc le changement de hauteur par une rotation appropriée de la tête de lecture.

Implémentation numérique

La plupart des modificateurs de pitch disponibles dans le commerce sont basés sur une réalisation numérique du système décrit ci-dessus. La bande magnétique est remplacée par une mémoire circulaire dans laquelle on place les échantillons du signal en entrée. Cette mémoire circulaire est lue par deux pointeurs diamétralement opposés (voir figure V.14

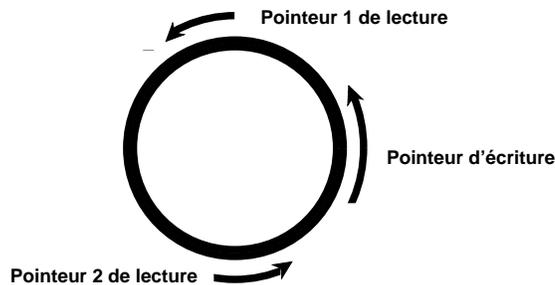


FIG. V.14 – *Système de lecture/écriture pour l'implémentation numérique de la technique de la mémoire circulaire*

A chaque échantillon écrit dans la mémoire (toutes les ΔT secondes), on avance les pointeurs de lectures de $\alpha \Delta T$ secondes, où α est le taux de modification, puis on lit un échantillon dans la mémoire. En général (pour les valeurs de α non entières), on se retrouve entre deux échantillons, et comme dans le cas du "flanger", il est nécessaire de calculer la valeur du signal à cet instant. Ici aussi, une simple interpolation linéaire convient.

Ainsi, le signal est lu avec une fréquence d'échantillonnage différente de celle à laquelle il a été écrit, ce qui provoque une modification de l'échelle des fréquences de taux α . Un problème se pose lorsque le pointeur de lecture rattrape (lorsque $\alpha > 1$) ou est rattrapé (lorsque $\alpha < 1$) par le pointeur d'écriture. Comme dans l'équivalent analogique, la continuité est assurée par un mixage des deux pointeurs au moment où se produit la rencontre ("cross-fade"): l'échantillon lu par le

pointeur de lecture courant (par exemple le pointeur 1) subit une pondération décroissante tandis que celui lu par l'autre pointeur de lecture (le pointeur 2) subit une pondération croissante. Finalement, le second pointeur devient le pointeur courant, et garde sa pondération maximale, jusqu'à ce que le pointeur d'écriture s'en rapproche.

Implémenté de cette façon, le pitch shifter a un comportement sensiblement équivalent à son homologue analogique (à ceci près qu'il est plus facilement paramétrable). Son implémentation en temps réel ne pose pas de problème particulier, puisqu'il ne réclame que très peu de calculs. Il produit malheureusement un bruit artificiel qui provient du mixage périodique des deux pointeurs de lecture. Pour tenter d'améliorer la qualité obtenue, on cherche à mieux raccorder les signaux lus par les deux pointeurs de lecture, de façon un peu similaire à ce qui est fait dans les méthodes synchrones. On peut par exemple utiliser la fonction d'autocorrélation du signal pour déterminer l'endroit le plus adéquat pour le "cross-fading" [6, 14].

Modification de la durée par la technique de la mémoire circulaire

Comme son homologue analogique, la technique du buffer circulaire peut être également utilisée pour la modification de l'échelle temporelle: Si l'on dispose d'un pitch shifter à mémoire circulaire, et que l'on veut effectuer un "time scaling" de paramètre α , il suffit de changer la fréquence d'échantillonnage du signal d'un taux α , puis de le traiter par le pitch shifter. Ainsi, pour ralentir le signal 2 fois, il suffit de le sur-échantillonner 2 fois. Si l'on écoute le signal obtenu à la fréquence originale, il sera deux fois plus long, mais aussi une octave plus grave. Il suffit donc de l'écouter à la fréquence originale en intercalant un pitch shifter de taux $\alpha = 2$.

On se rend rapidement compte qu'il est plus simple de faire les deux opérations une seule fois: La technique consiste alors à répéter ou éliminer périodiquement des portions de signal pour en augmenter (ou diminuer) la durée. Vue sous cet angle, cette technique (qui prend alors le nom de 'splicing method') se rapproche d'une technique TD-PSOLA dans laquelle on ne connaîtrait pas la valeur de la fréquence fondamentale. Les artefacts inhérents à cette méthode, qui proviennent des ruptures de la périodicité du signal lors des répétitions ou des éliminations peuvent être considérablement réduits par l'utilisation de méthodes basées sur l'autocorrélation du signal pour optimiser la longueur et l'emplacement des portions de signal à dupliquer ou à détruire [15, 6, 14, 20, 31, 32].

V.4.3 Harmoniseur

L'harmoniseur est une généralisation du changeur de fréquence (proposée la première fois dans les années 70 par Tony Agnello qui travaillait chez Eventide). Plusieurs changements de fréquence sont effectués simultanément, et leur sorties sommées. On choisit en général des changements de fréquence correspondant à la gamme tempérée (pour les applications musicales). Il s'agit alors d'une transposition.

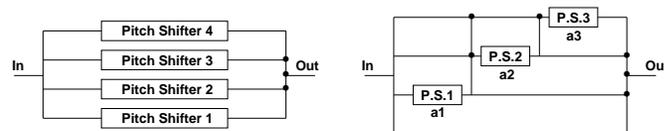


FIG. V.15 – Schéma d'un harmoniseur

La première implémentation (graphique de gauche sur la figure V.15) produit à partir d'un signal quatre transpositions. La seconde permet de produire plus de transpositions en économisant le nombre de transposeurs. A partir de trois transposeurs de rapport a_1 , a_2 et a_3 , le signal obtenu en sortie comporte les transpositions 1, a_1 , a_2 , a_3 , $a_1 \times a_2$, $a_1 \times a_3$, $a_2 \times a_3$, $a_1 \times a_2 \times a_3$. Avec N transposeurs on peut ainsi réaliser 2^N transpositions. Les harmoniseurs de ce type sont le plus souvent utilisés pour des transpositions d'octave car la transposition systématique d'un intervalle constant ne satisfait pas les règles les plus élémentaires de l'harmonie en musique (sauf si cet intervalle est une octave). Il suffit pour s'en convaincre d'observer la transposition à la tierce majeure d'un mode de Do majeur: le mode transposé n'appartient plus au mode de Do majeur puisqu'il inclut des notes "étrangères" (Fa dièse, Sol dièse, Do dièse et Ré dièse). Une nette amélioration des harmoniseurs est apparu récemment qui permet de mieux respecter les principes (élémentaires) de l'harmonie musicale: dans ce type d'harmoniseur, la transposition n'est pas fixe, mais varie en fonction de la note jouée de façon que la mélodie transposée appartienne à un mode musical donné. Dans l'exemple précédent, la transposition sera d'une tierce majeure ou mineure en fonction de la note jouée. Ce type d'harmoniseur utilise un détecteur de fondamental pour déterminer la fréquence du son et la note correspondante.

V.5 Le "chorusing"

A partir d'une source sonore unique, l'effet de "chorusing" cherche à imiter le son produit par l'écoute simultanée de plusieurs sources similaires. Ainsi pour une guitare, l'effet de chorusing tente de donner l'impression que plusieurs guitaristes jouent simultanément.

Les effets de "chorusing" reposent essentiellement sur la même idée: dans un ensemble de chanteurs ou d'instrumentistes, il n'y en a pas deux qui soient "accordés" de façon exactement similaire (c'est à dire, qui produisent exactement le même son, à la même hauteur). De plus, leur différence de hauteur ne reste pas constante dans le temps. D'où l'idée d'ajouter au signal initial des copies de hauteur légèrement différente, avec un désaccord variant dans le temps. Deux solutions peuvent être adoptées: l'utilisation de "pitch shifters" et l'utilisation de filtres passe-tout variant dans le temps.

V.5.1 Implémentation par changement de fréquence

Pour obtenir plusieurs copies du signal de hauteur légèrement différente, on peut utiliser un harmoniseur. En utilisant deux modules de changement de hauteur de rapports a_1 et a_2 , on peut obtenir les transpositions: 1, a_1 , a_2 et $a_1 \times a_2$. On a ainsi l'impression de quatre sources simultanées.

Les rapport a_1 et a_2 sont généralement très voisins de l'unité. En effet, si dans un groupe instrumental l'un des musiciens jouait plus d'un quart de ton en dessous des autres, le résultat serait musicalement inacceptable. En pratique, on utilise des écarts inférieurs à 2 commas (en musique, le comma est le neuvième du ton, soit 1.2%). De plus, on fait varier ces écarts dans le temps, pour éviter les battements audibles, etc...

Cette solution est toutefois relativement chère, puisqu'elle nécessite l'utilisation d'harmoniseurs.

V.5.2 Implémentation par filtre passe-tout à retard variable

Effet d'un simple retard variable

Lorsqu'un signal est lu avec un retard variable dans le temps, selon la figure ci-dessous, il subit en fait une opération de ré-échantillonnage.



FIG. V.16 – Lecture d'un signal avec un retard variable au cours du temps

En effet, en supposant le signal $x(t)$ connu à tous les instants t ,

$$y(t) = x(t - \tau(t)) \quad \text{devient} \quad y_n = y(n\Delta T) = x(n\Delta T - \tau(n\Delta T))$$

où l'on voit que le signal y_n correspond au signal $x(t)$ échantillonné aux instants $n\Delta T - \tau(n\Delta T)$. Si par exemple

$$\tau(t) = \alpha t \quad \text{alors} \quad y_n = x(n\Delta T(1 - \alpha))$$

ce qui montre que le signal y_n est obtenu par ré-échantillonnage du signal x_n à la nouvelle fréquence d'échantillonnage $Fe' = Fe/(1 - \alpha)$. Plus généralement, pour un retard $\tau(t)$ non linéaire en t , le signal est ré-échantillonné à une fréquence d'échantillonnage *locale*

$$Fe'(t) = Fe \left(1 - \frac{d\tau}{dt}\right)^{-1}$$

d'où l'idée d'utiliser un retard variable pour modifier localement la fréquence d'échantillonnage [3]. Ce ré-échantillonnage pose quelques problèmes de repliement spectral (voir plus loin).

Bien sûr, un retard linéaire en t ne convient pas car la ligne à retard a une durée finie. En pratique, on utilise un retard périodique, sinusoïdal ou en dents de scie, de fréquence ω_0 :

$$\tau(t) = \tau_0 + \alpha \sin(\omega_0 t) \quad \text{alors} \quad Fe'(t) = Fe (1 - \alpha\omega_0 \cos(\omega_0 t))^{-1} \approx Fe (1 + \alpha\omega_0 \cos(\omega_0 t))$$

où l'on voit que $Fe'(t)$ est tantôt supérieure tantôt inférieure à Fe , et la mémoire tampon a une longueur $\tau_0 + \alpha$. On peut également utiliser un retard à caractère aléatoire [3]. Puisque le signal y_n est calculé en échantillonnant $x(t)$ avec une fréquence d'échantillonnage localement variable, mais écouté à une fréquence d'échantillonnage constante, le signal original $x(t)$ subit une contraction/dilatation périodique de l'axe des fréquences, ce qui correspond bien à une variation de hauteur périodique.

Ce principe est à la base de l'implémentation de "chorus" en filtre passe-tout à retard variable.

Implémentation en passe-tout

Si la sortie du dispositif ci-dessus est simplement ajoutée au signal direct, on retrouve le schéma du "flanger", et de façon comparable, le signal subit de profondes altérations de son spectre. Pour éviter ce problème, on applique le principe du retard variable à un (ou plusieurs) filtre passe-tout (voir figure (V.17)) :

Ainsi, pour un retard donné (entier) p , la fonction de transfert s'écrit

$$H(z) = \frac{g + z^{-p}}{1 + gz^{-p}}$$

qui est bien celle d'un filtre passe-tout. Ainsi, on conserve la propriété de changement de fréquence périodique, en évitant les problèmes d'altération du spectre (en réalité, si le filtre *statique* est bien passe-tout, lorsque le retard varie la propriété passe-tout n'a plus vraiment de

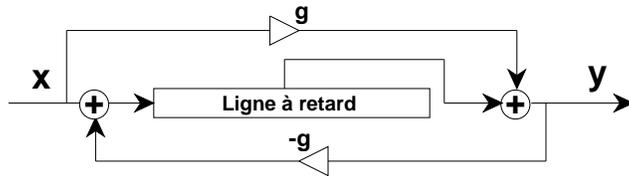


FIG. V.17 – Implémentation du chorusing par filtre passe-tout à retard variable

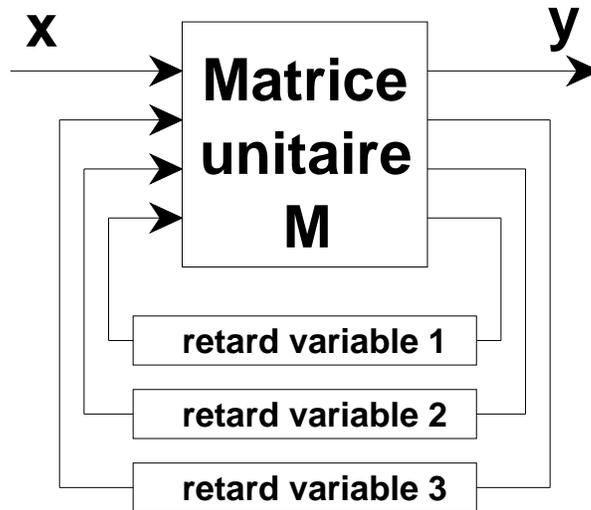


FIG. V.18 – Généralisation pour l'implémentation du chorusing par filtre passe-tout à retard variable

sens, et en pratique, on constate de légères variations timbrales).

Le schéma de la figure V.17 peut être généralisé à plusieurs lignes à retard de la façon suivante [18] (voir figure V.18:

On peut en effet montrer que si la matrice M est unitaire, alors le filtre réalisé par la figure ci-dessus (lorsque les retards sont fixes) est bien passe-tout. En faisant varier les retards de façon indépendante, on obtient plusieurs répliques du signal original avec des glissements de fréquence variables dans le temps.

Problèmes liés à l'interpolation linéaire

Comme pour l'effet de phasing, le retard doit varier dans le temps et pour éviter les problèmes de discontinuités en sortie, il est nécessaire d'utiliser des retards non-entiers. Lorsque le retard varie rapidement dans le temps, l'interpolation linéaire se traduit par un rééchantillonnage du signal qui n'est pas à bande limité! En effet, interpoler linéairement entre les échantillons du signal revient à convoluer le signal échantillonné par une fonction triangulaire de hauteur 1 et

de largeur deux périodes d'échantillonnage.

$$\tilde{x}(t) = \left(\sum_{n=-\infty}^{\infty} x_n \delta(t-n) \right) \star w(t) \quad \text{avec} \quad w(t) = \begin{cases} \Delta T(t + 1/\Delta T) & \text{pour } -1/\Delta T < t < 0 \\ \Delta T(1/\Delta T - t) & \text{pour } 0 < t < 1/\Delta T \end{cases}$$

Cette opération donne en sortie un signal à temps continu $\tilde{x}(t)$ dont le spectre est le spectre du signal original périodisé (à cause de l'échantillonnage), et multiplié par la transformée de Fourier de la fonction triangulaire, un sinus cardinal au carré. Le signal à temps continu $\tilde{x}(t)$ est ensuite ré-échantillonné avec une période d'échantillonnage $(1-\alpha)\Delta T$ et cela sans filtrage passe-bas préalable. Ce ré-échantillonnage qui n'est pas à bande limitée produit donc un repliement spectral. Pour les signaux audio dont le spectre est principalement localisé en basse-fréquence, ce repliement spectral ne pose pas de problème majeur [19]. De plus, l'effet de chorusing a tendance à masquer les artefacts dus au ré-échantillonnage imparfait.

V.6 Distorsion

V.6.1 Idée générale

Bien que la distorsion du signal soit un effet généralement indésirable dans une installation haute fidélité, c'est aussi un effet recherché dans certains cas, notamment par les guitaristes (encore eux!) Une distorsion du signal est aisément obtenue en le faisant circuler à travers une fonction de transfert non linéaire:

$$y(t) = f(x(t)) \quad \text{avec } f(u) \text{ non linéaire}$$

La fonction de transfert non linéaire fait apparaître dans le spectre des composantes qui n'y figuraient pas originellement. En particulier, si x_n est composé d'une somme de sinusoides, et si l'on peut développer $f(u)$ en série de Maclaurin autour de 0,

$$f(u) = f(0) + uf'(0) + \frac{u^2 f''(0)}{2!} + \frac{u^3 f'''(0)}{3!} + \dots = \sum_{i=0}^{\infty} \frac{u^i}{i!} \frac{d^i f}{du^i}$$

on voit que le signal x_n

$$x_n = \sum_{i=-p}^p e^{j\omega_i n}$$

passé à travers la fonction f apparaît comme une somme pondérée de ses puissances entières, et l'on a donc

$$y_n = \sum_{i=0}^{\infty} \alpha_i (x_n)^i \quad \text{où} \quad (x_n)^i = \sum_{-p \leq j(k) \leq p} e^{j(\omega_{j(0)} + \omega_{j(1)} + \dots + \omega_{j(i-1)})}$$

Pour un signal composé de sinusoides réelles, les termes ω_i apparaissent par paires de signes opposés. Par suite, le signal en sortie y_n contient des composantes sinusoidales dont les fréquences sont toutes les sommes et les différences des multiples des fréquences originales ω_i . Ainsi, un couple de sinusoides $x_n = \cos(\omega_1) + \cos(\omega_2)$ donnera en sortie un signal qui en général contiendra les fréquences $\omega_1, \omega_2, \omega_1 + \omega_2, \omega_1 - \omega_2, 2\omega_1$ etc...

Si x_n est périodique, y_n est également périodique de même période, mais son contenu spectral est modifié.

V.6.2 Implémentation à l'aide de polynômes de Chebyshev

Une façon de choisir la fonction de transfert consiste à considérer en entrée un signal purement sinusoïdal, et tenter de contrôler le nombre et les amplitudes des harmoniques qui apparaissent après transformation non-linéaire. On utilise pour cela les polynômes de Chebyshev T_i qui possèdent la propriété

$$T_i(\cos(\theta)) = \cos(i\theta)$$

Ces polynômes peuvent être calculés par:

$$T_0(x) = 1, T_1(x) = x \quad \text{et} \quad T_{n+1}(x) - 2xT_n(x) + T_{n-1}(x) = 0$$

Cette propriété intéressante permet de doser le degré de distorsion obtenu par le système. On voit qu'en choisissant pour $f(u)$ une somme pondérée de polynômes de Chebyshev, on fait apparaître dans le signal $f(\cos(\omega n))$ une somme pondérée de certaines fréquences multiples de ω . Ci-dessous, un exemple de fonction de transfert obtenue de cette façon, avec un ordre maximal de 3 ($P(x) = 0.12x^3 + 0.4x^2 - 0.91x - 0.2$), et un signal original comportant deux sinusoïdes avant et après distorsion.

où l'on voit que le signal traité comporte les fréquences multiples des deux fréquences originales (jusqu'à un facteur 3), mais aussi les multiples de leur différences etc...

Puisque la fonction $f(u)$ est non-linéaire, la distorsion générée par son application au signal x_n dépend fortement de l'amplitude maximale du signal: pour un signal de très faible niveau, $f(u)$ peut être considérée comme linéaire (si son terme constant est nul) et la distorsion devient faible.

V.7 Réverbération: introduction

V.7.1 Canal acoustique

Dans un milieu de propagation acoustique, le signal capté par un récepteur diffère du signal émis par une source sonore. Cette différence est due à des phénomènes physiques qui, en première approximation, peuvent être considérés comme linéaires et représentés par une réponse impulsionnelle $h(t)$ causale liant le signal émis $x(t)$ et le signal reçu $y(t)$:

$$y(t) = \int_0^\infty x(t-u)h(u)du \quad \text{ou} \quad y_n = \sum_{i=0}^\infty x_{n-i}h_i \quad (\text{V.1})$$

La transformation subie par le signal peut être modélisée comme une transmission dans ce qu'on appellera le canal acoustique [9].

V.7.2 Réverbération en lieu clos

Dans un lieu clos, la présence des parois et d'obstacles entraîne l'existence de trajets multiples de durées distinctes entre la source et le récepteur, ce qui se traduit, dans le signal reçu, par une "répétition" du signal émis due à la superposition des contributions des différents trajets, phénomène auquel on donne le nom de réverbération. La première des contributions possibles est le son direct, c'est-à-dire celui qui existerait en l'absence des parois et de tout autre obstacle. La somme des contributions de tous les autres trajets forme l'effet de salle. Ici, le canal acoustique est entièrement défini par un ensemble de paramètres que l'on conviendra de diviser en trois sous-ensembles:

- la salle, définie par la géométrie et les matériaux des murs et des obstacles,

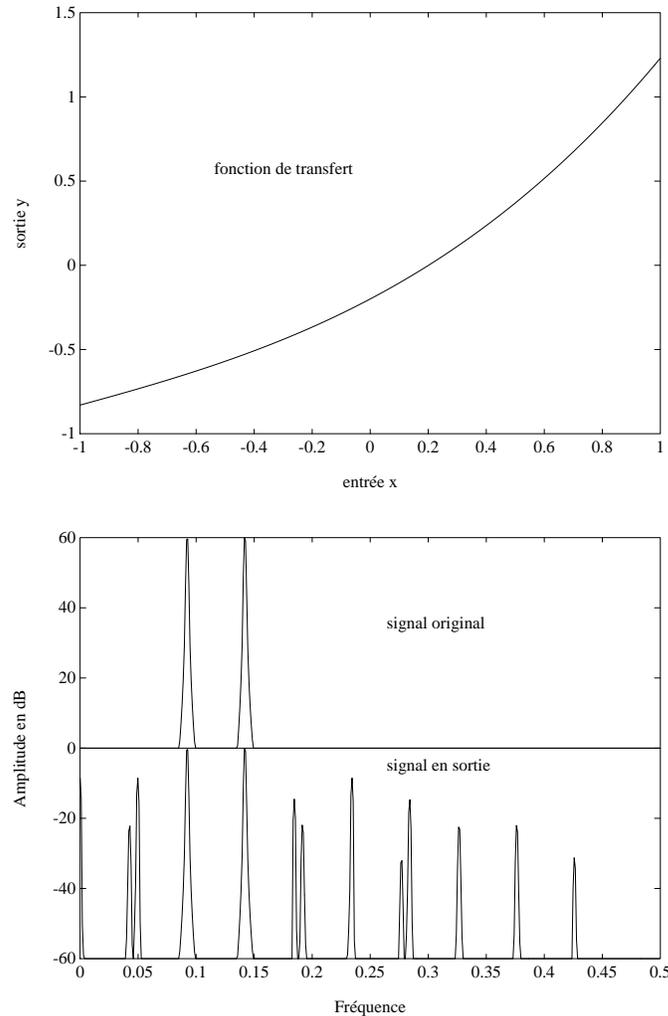


FIG. V.19 – Illustration de l'utilisation de polynômes de Chebyshev pour l'obtention de fonction non-linéaire pour obtenir un effet de distorsion

- la source, définie par sa position, son diagramme de rayonnement (directivité d'émission) et son étendue,
- le récepteur (typiquement un voire deux microphones), défini par sa position et son diagramme de directivité.

La figure V.20 présente l'échogramme-type d'une réponse impulsionnelle (monophonique) enregistrée dans une salle, où l'on distingue les contributions du son direct et de l'effet de salle. Dans ce dernier, on distingue encore les premières réflexions (ou échos précoces), temporellement séparables, d'un "continuum" perçu par l'oreille comme une traînée sonore. Cette traînée sonore, qu'on appellera réverbération tardive (expression française correspondant au terme anglais late reverberation) peut être considérée, sur le plan perceptif, comme caractéristique de la salle elle-même, c'est-à-dire indépendante des paramètres liés à la source et au récepteur, ce qui n'est pas le cas en revanche pour les premières réflexions [9]. Comme l'illustre la figure V.20, l'énergie de la réverbération tardive décroît suivant une allure approximativement exponentielle (donc linéairement lorsqu'elle est exprimée en décibels).

V.7.3 Auralisation

On regroupe aujourd’hui sous le vocable ”auralisation” les techniques qui visent à restituer auditivement un environnement sonore sur un dispositif de reproduction individuel. Ces techniques trouvent leur application pour l’aide à la construction de salles de concert, dans les studios d’enregistrement, mais aussi, associées à la vidéo ou à la synthèse d’image, dans les systèmes de simulation et de réalité virtuelle ou les futurs systèmes d’audioconférence. La relation V.1 et la connaissance de la réponse impulsionnelle h_n offrent un premier moyen de recréer artificiellement l’effet de réverbération quel que soit le signal émis x_n , au moyen d’un ordinateur numérique.

Le calcul de la convolution discrète d’après l’équation V.1 nécessite N additions-multiplications par échantillon si N est la durée de la réponse impulsionnelle exprimée en nombre d’échantillons. Pour une durée de une seconde et une fréquence d’échantillonnage de 32 kHz, ce coût dépasse d’un facteur 50 environ la puissance des microprocesseurs de traitement du signal les plus récents. Les techniques de convolution rapides fondées sur l’algorithme de FFT permettent de réduire ce coût à un ordre de grandeur raisonnable: environ $4\log_2(2N)$ opérations par échantillon pour la méthode ”overlap-save” avec FFT sur $2N$ points [4]. Cependant, cette méthode nécessite une capacité mémoire de $5N$ échantillons et introduit un retard de traitement d’une durée de $4N$ échantillons.

Outre l’important volume de calcul, le filtrage par convolution se révèle peu pratique car le contrôle de l’effet de réverbération artificielle nécessite la remise à jour globale de tous les coefficients constituant la réponse impulsionnelle, soit plusieurs milliers! On est donc naturellement conduit à rechercher des modèles, fondés sur des propriétés physiques et perceptives du phénomène de réverbération, permettant de simplifier la mise en oeuvre et le contrôle en temps réel du filtrage numérique simulant un canal acoustique.

V.7.4 Modèles physiques et perceptifs en acoustique des salles

Depuis les années 70, les recherches en acoustique des salles [25, 13, 1] ont permis de montrer que, tant du point de vue physique que du point de vue perceptif, il est justifié de décomposer l’effet de salle en ‘réverbération tardive’ d’une part (principalement fonction du volume et de l’absorption globale de la salle), et ‘réflexions précoces’ d’autre part (qui dépendent de la géométrie de la salle et de la position de la source et du récepteur) [13].

La réverbération tardive est caractérisée par une grande densité d’échos par unité de temps, de sorte qu’on peut la modéliser sous la forme d’un processus aléatoire gaussien décroissant exponentiellement au cours du temps. De même, la réponse en fréquence de la salle possède une densité de modes très élevée, du moins au delà d’une fréquence limite (souvent appelée fréquence de Schroeder) qui va de quelques dizaines de Hz (grandes salles) à quelques centaines de Hz (petites salles relativement réverbérantes). La figure V.20 (graphique du bas) présente le domaine de validité du modèle statistique en fonction du temps de mélange \sqrt{V} et de la fréquence de Schroeder $2\sqrt{\frac{Tr}{V}}$.

La caractérisation perceptive de la qualité acoustique distingue les aspects spatiaux (directionnels) des aspects temporels de l’effet de salle. On peut caractériser les attributs spatiaux par une mesure de similarité binaurale dérivée de la fonction d’intercorrélation entre les réponses impulsionnelles droite et gauche mesurées sur une tête artificielle [5].

Les attributs temporels, liés à la répartition de l’énergie dans la réponse impulsionnelle, sont mis en évidence sur la courbe de décroissance, qui décrit la puissance du signal reçu en fonction du temps écoulé après que la source a cessé d’émettre un message continu (le message de référence étant un bruit blanc). Schroeder a montré que la courbe de décroissance (Energy Decay Curve) se

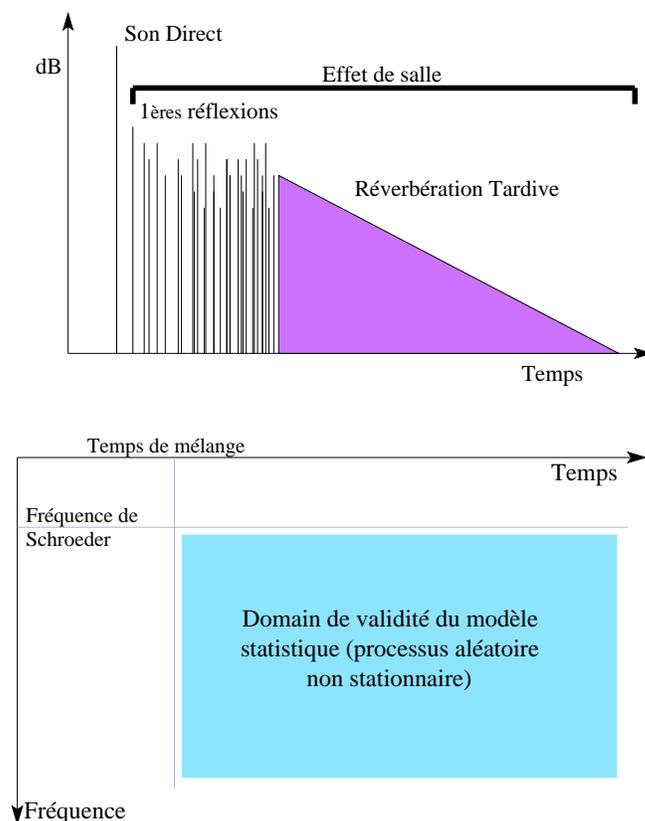


FIG. V.20 – Premières réflexions et réverbération tardive. Validité du modèle statistique en fonction du temps de mélange et de la fréquence de Schroeder

déduit directement de la réponse impulsionnelle $h(t)$ par intégration rétrograde de l'énergie [24]:

$$EDC_h(t) = \int_t^{+\infty} h^2(\tau) d\tau \quad (\text{V.2})$$

$EDC(t)$ est donc l'énergie restant dans la réponse impulsionnelle après l'instant t , habituellement représentée en décibels (voir fig. V.20). Le principal critère utilisé par les acousticiens des salles, le temps de réverbération, mesure la pente de la courbe de décroissance (exprimée par le temps nécessaire pour que l'énergie décroisse de 60 dB). Le temps de réverbération varie en fonction de la fréquence ce qui nécessite un calcul de la courbe de décroissance par bandes d'octaves, à partir de versions filtrées de la réponse $h(t)$. Une procédure équivalente consiste à représenter, en fonction du temps t , la répartition fréquentielle de l'énergie restant dans la réponse impulsionnelle après l'instant t . On définit ainsi le "relief de décroissance" ("Energy Decay Relief") [11, 12].

V.8 Algorithmes de réverbération artificielle

Les réverbérateurs artificiels sont utilisés dans l'industrie musicale ou cinématographique pour ajouter un "effet de salle" aux enregistrements effectués en studio, ou encore pour modifier les propriétés acoustiques d'une salle d'écoute. Les premiers appareils de ce type, apparus dans

les années 50, fabriquaient des échos artificiels par des procédés analogiques: réseaux de ressorts, puis plaques métalliques, munis de transducteurs électromagnétiques. Avec les magnétophones apparurent les effets utilisant des bandes magnétiques en boucle, où un retard était naturellement introduit par la distance entre la tête d'enregistrement et la tête de lecture, et sa durée réglable par la vitesse de défilement [10].

Vers la fin des années 70, les convertisseurs analogique-numérique offrirent la possibilité de réaliser des unités de retard peu coûteuses, ayant une faible distorsion et un excellent rapport signal/bruit. La réverbération artificielle numérique fut alors une des premières applications en temps réel des technologies audio-numériques naissantes. Sa faisabilité avait été démontrée dès le début des années 60 par Schroeder [22], dont les travaux ont inspiré, dans la dernière décennie, un grand nombre de réalisations commerciales [10].

Pour reproduire en temps réel un effet de salle tardif sans recourir à la convolution du signal traité par une réponse impulsionnelle mesurée ou modélisée, il est nécessaire de concevoir un filtre numérique récuratif ("filtre réverbérant") dont la réponse impulsionnelle satisfasse les contraintes du modèle statistique de la réverbération tardive décrit au paragraphe V.7.4. Les difficultés rencontrées dans la réalisation de filtres réverbérants sont de trois natures (voir par exemple [17] ou [10]):

1. Obtenir simultanément une densité modale et une densité d'échos suffisantes,
2. éviter l'apparition, dans le signal de sortie, d'une sonorité peu naturelle, souvent qualifiée de "métallique", notamment lorsque le signal d'entrée est de type percussif,
3. contrôler indépendamment le temps de réverbération et l'énergie réverbérée en fonction de la fréquence.

V.8.1 Filtres en peigne et filtres passe-tout

Pour réaliser des réverbérateurs artificiels, Schroeder [23] proposa au début des années 60 une approche, approfondie plus tard par Moorer [17], fondée sur l'association de filtres en peignes et de filtres passe-tout. On peut penser que ces filtres sont les constituants de base de la plupart des réverbérateurs du commerce actuels [21, 27, 10].

Les fonctions de transfert du filtre en peigne et du filtre passe-tout représentés sur la figures V.21 et V.22 sont respectivement:

$$C(z) = \frac{z^{-m}}{1 - gz^{-m}} \quad \text{et} \quad A(z) = \frac{-g + z^{-m}}{1 - gz^{-m}}$$

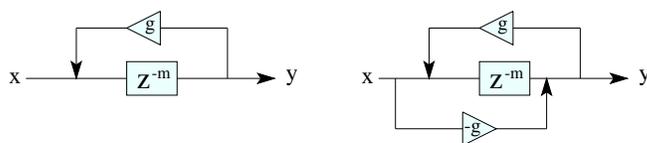


FIG. V.21 – *Filtre en peigne récuratif (à gauche) et filtre passe-tout (à droite).*

Le filtre en peigne imprime au signal traité une coloration harmonique très caractéristique car ses résonances sont situées à des fréquences harmoniques. Sa réponse en fréquence est maximale pour la série de fréquences $\omega_k = \frac{2\pi k}{m}$ si $g > 0$ ou $\omega_k = \frac{(1+2k)\pi}{m}$ si $g < 0$. Les résonances correspondantes ont toutes le même taux d'amortissement.

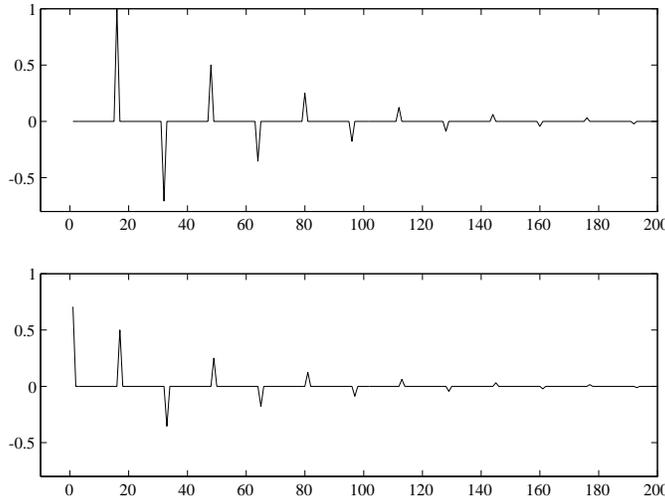


FIG. V.22 – Réponse impulsionnelle du filtre en peigne et du filtre passe-tout ($m = 16, g = -0.707$).

Schroeder a montré que la simple addition d'un trajet direct, d'amplitude bien choisie, suffit à transformer le filtre en peigne en un filtre passe-tout, donc à supprimer la coloration harmonique. Néanmoins, le filtre passe-tout engendre le même défaut de timbre que le filtre en peigne lorsque le signal d'entrée n'est pas stationnaire, ce qui est le cas des signaux audio de parole et de musique. Dans le domaine temporel, les inconvénients majeurs de ces deux filtres résultent de la faible densité d'échos de leurs réponses impulsionnelles (fig. V.21 et V.22), qui se traduit par des "roulements" (traduction du terme anglais flutter) dans la réponse à des signaux transitoires.

L'association en série de filtres passe-tout (fig. V.23) permet d'augmenter la densité d'échos tout en conservant le caractère passe-tout du réverbérateur. Cependant, le filtre "passe-tout série" possède, en présence de signaux transitoires, une sonorité métallique peu naturelle [17, 26]. Notons, qu'il existe d'autres façons de réaliser des filtres passe-tout, en particulier à l'aide de la structure treillis.

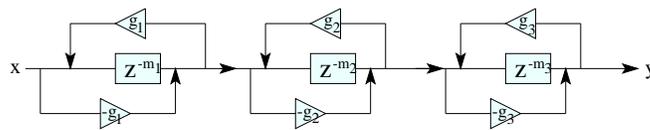


FIG. V.23 – Structure passe-tout série proposée par Schroeder.

V.8.2 Le réverbérateur de Schroeder

Afin d'éliminer toute coloration, tout en améliorant la densité d'échos, Schroeder [23] a proposé d'utiliser en cascade deux combinaisons de base à partir des filtres précédents: une association série de filtres passe-tout et une association parallèle de filtres en peigne (fig. V.24). L'association en parallèle de filtres en peigne ne permet pas d'obtenir une réponse en fréquence plate. Néanmoins, cette réponse devient comparable à celle d'une salle réelle à condition de présenter une densité modale (nombre de modes propres par Hz) suffisamment élevée.

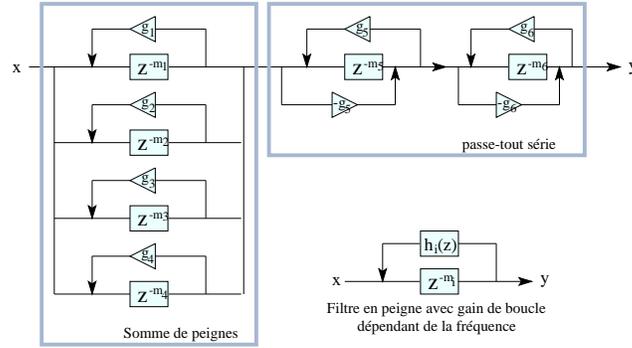


FIG. V.24 – *Filtre réverbérant à base de filtres en peigne et de filtres passe-tout proposé par Schroeder. En bas à droite, remplacer le gain g_i par un filtre permet de simuler une absorption dépendant de la fréquence.*

Pour un filtre en peigne, le temps de réverbération Tr est donné par:

$$\frac{20 \log_{10}(g_i)}{m_i T} = \frac{-60}{Tr} \quad (V.3)$$

où g_i est le coefficient (ou gain) de bouclage, m_i la durée du retard (en nombre de périodes d'échantillonnage) de la cellule i , et T est la période d'échantillonnage. Pour la somme de peignes de la fig. V.24, Schroeder propose d'affecter à chaque filtre en peigne la même valeur de Tr , de sorte que le choix du gain de boucle g_i dépend de la durée du retard m_i . Le respect de cette condition est indispensable pour éviter que la sonorité caractéristique de l'un des filtres en peignes ne soit audible dans la fin de la réponse à des sons transitoires. Schroeder conseille de choisir les durées m_i des retards premières entre elles pour éviter les superpositions et annulations d'échos, ce qui a aussi pour effet de rendre les fréquences de résonance distinctes deux à deux, et de rompre toute périodicité dans la réponse fréquentielle.

Dans le cas de N filtres en peigne en parallèle, la densité modale et la densité d'échos s'obtiennent en sommant celles des filtres élémentaires, sachant que, pour chaque filtre en peigne, les échos successifs sont séparés d'un intervalle de temps égal à la durée du retard $\tau_i = m_i T$ et les fréquences propres voisines sont distantes d'un intervalle de fréquence égal à $1/\tau_i$:

$$\text{densité modale: } D_m = \sum_{i=0}^{N-1} \tau_i = N\tau, \text{ densité d'échos: } D_e = \sum_{i=0}^{N-1} \frac{1}{\tau_i} \approx \frac{N}{\tau}$$

où τ est la durée moyenne des retards τ_i . Ces expressions appellent deux remarques: la première est que ces densités sont indépendantes du temps ou de la fréquence, alors que la théorie modale et la théorie géométrique de la propagation du son dans une salle indiquent que D_m est proportionnelle au carré de la fréquence et D_e proportionnelle au carré du temps écouté dans la réponse impulsionnelle. Cela ne pose pas de problème, dans le cadre du modèle statistique de la réverbération tardive, à condition que ces deux densités soient suffisamment élevées dans le filtre réverbérant. La seconde remarque est que, pour un nombre N donné, le choix des durées τ_i procède d'un compromis entre densité modale et densité d'échos. A condition que ces durées restent suffisamment voisines pour que l'expression de D_e ci-dessus soit valable, on peut prédire le nombre N de filtres en peigne nécessaires en fonction des critères D_m et D_e :

$$N \approx \sqrt{D_m D_e}$$

Il reste alors à déterminer pour les densités D_m et D_e des valeurs qui soient perceptivement suffisantes. Pour juger de la densité d'échos, il est naturel de choisir l'impulsion élémentaire comme signal de test, révélateur du comportement du filtre réverbérant en présence de signaux percussifs. Selon Griesinger [10], la densité d'échos nécessaire pour un réverbérateur artificiel est de l'ordre de 10000 échos/s, et croît en fonction de la fréquence d'échantillonnage. Pour fixer la densité modale minimale nécessaire, Schroeder [23] se base sur le nombre de maxima par Hertz dans le modèle statistique de la réponse fréquentielle d'une salle, ce qui conduit à une densité modale approximativement égale au quart du temps de réverbération [22]. Pour un temps de réverbération de l'ordre de la seconde, on constate donc qu'il faudrait associer environ 50 filtres en peigne en parallèle! La solution proposée par Schroeder pour conserver un nombre raisonnable de cellules élémentaires est une structure mixte associant en cascade la somme de peignes avec un filtre passe-tout série (fig. V.24), dont le rôle est d'augmenter la densité d'échos sans modifier le timbre de la réverbération tandis que les filtres en peigne, dont le nombre peut ainsi être réduit à 4 ou 6 [22, 17], déterminent le temps de réverbération et la densité modale. Néanmoins, le choix du nombre de filtres passe-tout et de leurs paramètres (gains et retards) est très délicat et reste essentiellement empirique [23, 17].

V.8.3 Simulation d'un temps de réverbération dépendant de la fréquence

Pour réaliser un dispositif capable de fournir un temps de réverbération fonction de la fréquence, Schroeder a proposé de rendre les gains de boucle des filtres en peigne dépendants de la fréquence [23]. Cette idée a été mise en pratique par Moorer [17] dans le cadre du réverbérateur de la fig. V.24. Ce dernier propose de remplacer chaque atténuation g_i par un filtre passe-bas purement récursif du premier ordre, noté $h_i(z)$ (fig. V.24), en tentant de simuler aussi fidèlement que possible l'absorption dans l'air (l'absorption du son par les parois de la salle peut être négligée dans des salles de grandes dimensions). Cette méthode ne permet pas de prévoir ni de contrôler explicitement la variation résultante du temps de réverbération en fonction de la fréquence.

Le fait de remplacer les gains de boucle par des filtres passe-bas implique que l'énergie du signal réverbéré est diminuée dans les hautes fréquences. Si l'on souhaite assurer l'indépendance des réglages du temps de réverbération et de l'énergie du signal réverbéré en fonction de la fréquence, on peut associer en série avec le réverbérateur un filtre correcteur dont la réponse en fréquence énergétique doit être inversement proportionnelle au temps de réverbération [11, 12].

V.8.4 Systèmes unitaires

Gerzon [8] s'est intéressé à la réalisation d'un réverbérateur artificiel à plusieurs canaux d'entrée et de sortie (système caractérisé par sa réponse en fréquence matricielle $A(f)$ sur la fig. V.25. Il montre qu'un tel système préserve l'énergie de tout signal d'entrée si et seulement si, quelle que soit la fréquence f , la matrice $A(f)$ est unitaire (c'est-à-dire que son inverse est égale à sa transposée). Il appelle système unitaire tout système vérifiant cette propriété. Cette définition se transpose sans difficulté au cas des signaux discrets: un système de matrice de transfert $A(z)$ sera dit unitaire si et seulement si la matrice $A(z)$ est unitaire pour tout nombre complexe z de module 1.

La solution triviale pour réaliser des systèmes unitaires complexes à partir de systèmes unitaires simples consiste à associer des systèmes unitaires en cascade. Gerzon propose une seconde structure, représentée sur la fig. V.25, qui n'est autre qu'une généralisation du filtre passe-tout proposé par Schroeder. Pour montrer que cette structure forme un système unitaire, Gerzon fait appel à l'algèbre des matrices normales [8]. Afin d'obtenir un temps de réverbération

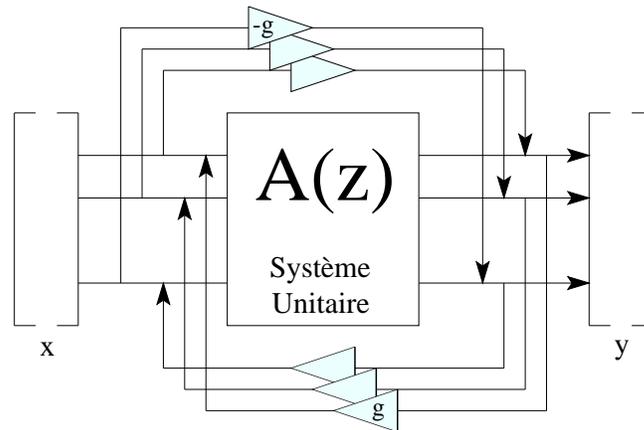
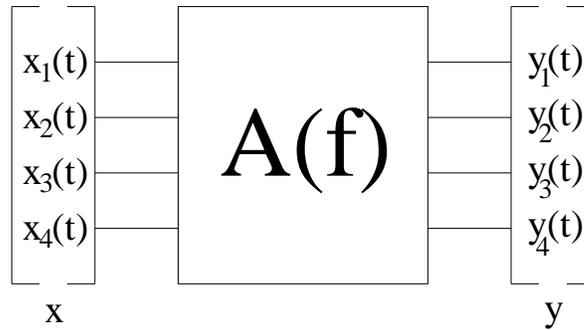


FIG. V.25 – *Système linéaire aux signaux d’entrée et de sortie vectoriels (à gauche); Structure proposée par Gerzon formant un système unitaire si le système de matrice de transfert $A(z)$ est lui-même unitaire (à droite)*

dépendant de la fréquence, Gerzon montre que le gain g peut être remplacé par un filtre de fonction de transfert $g(z)$, sans perte du caractère unitaire du système à condition que $g(z)$ soit à phase minimale.

V.8.5 Filtres réverbérants à bouclage unitaire

En imposant qu’un filtre réverbérant soit passe-tout (ou unitaire), on ne garantit pas l’absence de colorations métalliques en présence de signaux transitoires. Pour obtenir un effet de réverbération naturel, il est plus important d’assurer une densité modale suffisamment élevée que de préserver le spectre du signal d’entrée. Nous avons vu pour le filtre "somme de peignes" que la densité modale (en nombre de modes par Hz) est égale à la durée totale des retards exprimée en secondes et que la principale limitation réside dans la difficulté d’obtenir une densité d’échos suffisante avec un nombre N réduit de filtres élémentaires.

Une approche permettant d’augmenter le nombre d’échos dans la réponse du filtre réverbérant sans augmenter le nombre de retards consiste à associer des filtres élémentaires en série, ou bien, plus généralement, à ajouter des canaux de bouclage reliant les différents retards. L’effet de ce bouclage multiple est que la densité d’échos croît progressivement dans la réponse impulsion-

nelle, comme c'est le cas dans les salles réelles. Ce type de bouclage pose cependant un nouveau problème: celui de la stabilité du système résultant.

Stautner et Puckette [30] ont adopté une approche générale des réverbérateurs multi-canaux en proposant la structure représentée sur la fig. V.26, qui fait apparaître une matrice carrée A regroupant les coefficients de bouclage a_{ij} . Une approche équivalente, quoique utilisant un formalisme différent (celui des guides d'ondes), a été proposée par Smith [28]. On peut remarquer que la structure de la fig. V.26 est au filtre en peigne ce que les systèmes unitaires multi-canaux de Gerzon sont au filtre passe-tout. Stautner et Puckette montrent que la stabilité du système est garantie si la matrice de bouclage A est le produit d'une matrice unitaire et d'une matrice diagonale dont tous les coefficients diagonaux sont de module inférieur à 1. Ils proposent une application particulière à quatre canaux utilisant la matrice A suivante, unitaire au coefficient g près, qui sert à faire varier le temps de réverbération:

$$A = \frac{g}{\sqrt{2}} \begin{bmatrix} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \quad \text{avec } g < 1 \quad (\text{V.4})$$

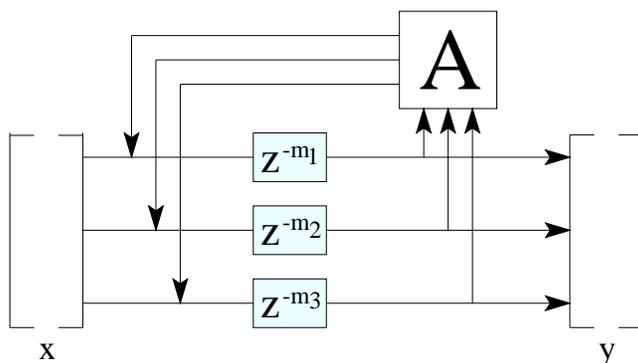


FIG. V.26 – *Modèle général de bouclage proposé par Stautner et Puckette.*

Dans cette application, la densité d'échos résultante n'est pas maximale car un coefficient de bouclage sur deux est nul. Le contrôle du temps de réverbération reste heuristique et inspiré des travaux de Moorer. Afin de réduire le temps de réverbération en hautes fréquences, un filtre passe-bas est associé à chaque retard. Pour remédier aux résonances métalliques qu'ils constatent, Stautner et Puckette proposent d'introduire une variation aléatoire des durées des retards au cours du temps. Cette technique est aussi utilisée par Smith [28] et dans les réverbérateurs de marque Lexicon [10], mais complique la réalisation pratique du réverbérateur.

On peut remarquer que le fait d'utiliser l'unique paramètre g pour contrôler le temps de réverbération revient à affecter une atténuation commune à tous les retards sans tenir compte de leurs durées. Dans le cas où la matrice A est diagonale, qui correspond au filtre somme de peignes étudié au paragraphe V.8.1, cela confère à la réverbération artificielle une sonorité métallique car les résonances n'ont pas toutes le même taux d'amortissement. Pour éviter ce phénomène, il faut faire en sorte que tous les pôles du filtre réverbérant aient le même module, ce qui peut être réalisé quelle que soit la matrice de bouclage A grâce aux propriétés ci-dessous [11].

On peut montrer que les pôles d'un système unitaire bouclé sur lui-même sont tous situés sur le cercle unité. Cela signifie que la réponse impulsionnelle est une somme de sinusôdes non amorties, c'est à dire que le réverbérateur possède un temps de réverbération infini. C'est le cas pour le système de la fig. V.26 si la matrice de bouclage est unitaire.

On montre que le fait d'affecter à chaque retard m_i une absorption g_i qui, exprimée en décibels, est proportionnelle à la durée m_i , suivant la relation V.3, a pour effet de multiplier la réponse impulsionnelle par une enveloppe temporelle exponentiellement décroissante réalisant le temps de réverbération désiré. Pour réaliser un contrôle explicite du temps de réverbération en fonction de la fréquence, il suffit de généraliser la méthode précédente en affectant à chaque retard un filtre absorbant dont la fonction de transfert $h_i(z)$ vérifie:

$$20 \log_{10} |h_i(e^{j\omega})| = \frac{-60T}{Tr(\omega)} m_i$$

Cette relation peut être satisfaite approximativement pour des filtres absorbants récursifs du premier ordre [11], ce qui offre deux paramètres de contrôle permettant de régler le temps de réverbération en basses fréquences et en hautes fréquences séparément. Un contrôle plus fin du temps de réverbération peut être réalisé à l'aide de filtres absorbants plus complexes, afin par exemple de reproduire la réverbération tardive d'une salle existante [12].

Cette méthode de contrôle de la décroissance de la réverbération est applicable quelle que soit la structure du bouclage du filtre réverbérant. Dans le cas le plus général, cette structure est un réseau de retard récursif qui doit être choisi, en fonction du nombre de canaux désiré en entrée et en sortie du système de façon à maximiser le nombre de trajets de bouclage entre les retards. La durée totale des retards doit être suffisamment élevée pour assurer une densité modale suffisante, et leur nombre suffisamment grand pour assurer une densité d'échos suffisante. En pratique, 12 retards d'une durée totale de 1 seconde suffisent pour reproduire une réverbération tardive naturelle. Lorsque les contraintes sur la densité modale et la densité d'échos sont satisfaites, le filtre réverbérant peut être comparé à un générateur de bruit blanc fournissant sur ses divers canaux de sortie des réalisations décorréélées d'un même processus aléatoire. Cependant, pour simuler de manière convaincante la réverbération d'une salle réelle, il convient d'apporter un soin particulier à la reproduction des réflexions précoces.

V.8.6 Contrôle des réflexions précoces

La plupart des réverbérateurs de haut de gamme permettent d'ajouter à la réverbération artificielle un groupe d'échos précoces. Ces échos jouent un rôle important dans le réalisme de la réverbération et influent en particulier sur les aspects perceptifs de clarté du message sonore (liée au rapport entre l'énergie précoce et l'énergie tardive de l'effet de salle) et d'impression d'espace [13]. Dans les approches proposées à ce jour pour contrôler les premiers échos produits par un réverbérateur artificiel, la distribution des dates d'arrivées et des amplitudes des premières réflexions est spécifiée à partir d'une modélisation géométrique de la propagation acoustique dans la salle, pour une disposition donnée de la source et du récepteur au sein de celle-ci. Cela suppose que l'on néglige les phénomènes physiques de diffraction et de diffusion, ce qui reste à justifier par des expériences psychoacoustiques.

Schroeder [25] propose d'ajouter les échos précoces au signal de sortie du filtre réverbérant, au moyen d'un filtre FIR permettant le contrôle de leurs amplitudes b_i et de leurs instants d'arrivée t_i par rapport au son direct (fig. V.27). Dans le réverbérateur de Moorer [17], les échos précoces contribuent aussi au signal d'entrée du réverbérateur, le gain r permettant le contrôle de la clarté (fig. V.29). Cette modification permet d'accroître la densité d'échos produite par

le filtre réverbérant, mais l'insertion du filtrage FIR rend le timbre de la réverbération tardive extrêmement tributaire de la distribution des premières réflexions. Ce principe peut provoquer une coloration non désirée de la réverbération et nécessite un soin particulier dans le choix de la distribution des premiers échos, comme le remarque Moorer [17]. Cette précaution est incompatible avec un calcul automatique des réflexions par simulation géométrique. Stautner et Puckette proposent un réverbérateur pouvant recevoir un nombre quelconque de signaux sources en entrée, et possédant une réponse distincte pour chacune en fonction de sa position dans la salle (fig. V.27 et V.28). Le signal provenant de chaque source vient s'ajouter en plusieurs points de chaque ligne à retard du filtre réverbérant afin de créer autant d'échos précoces. Les quatre signaux de sortie sont dirigés vers des haut-parleurs disposés autour de l'auditeur, ce qui permet un contrôle approximatif des directions de provenance des échos précoces. Cependant, le problème de coloration rencontré dans le système de Moorer se retrouve, dans une moindre mesure, dans celui de Stautner et Puckette. D'autre part, contrairement aux deux précédents, ce système ne permet pas de contrôler la clarté: l'énergie de la réverbération tardive est déterminée par celle des premiers échos.

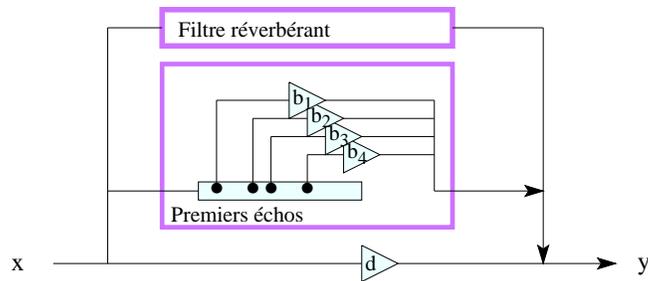


FIG. V.27 – *Approche proposée par Schroeder pour le contrôle des premiers échos.*

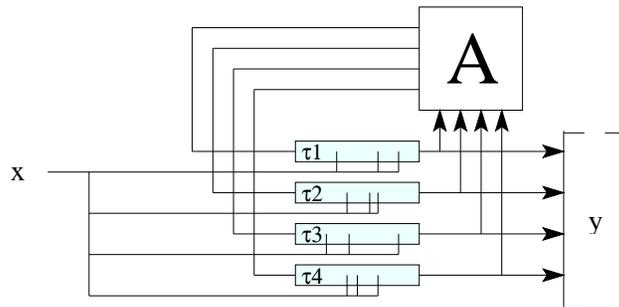


FIG. V.28 – *Approche proposée par Stautner et Puckette..*

La fig. V.30 décrit une variante du principe proposé par Stautner et Puckette, permettant de contrôler le niveau des premiers échos indépendamment de celui de la réverbération tardive. La ligne à retard qui permet le contrôle des dates t_i des échos précoces fournit une seule réplique du signal source sur chaque canal d'entrée du filtre réverbérant. Ce processus étant unitaire, le spectre de la réverbération tardive n'est pas affecté par le choix de la distribution temporelle des échos précoces. Ce principe de reproduction des aspects temporels peut être généralisé pour la spatialisation simultanée (mixage) de plusieurs sources [11] et appliqué indépendamment du choix du dispositif d'écoute (écouteurs ou haut-parleurs), qui détermine essentiellement la

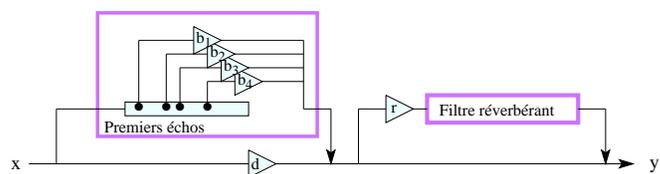


FIG. V.29 – Approche proposée par Moorer pour le contrôle des premiers échos.

restitution des informations directionnelles du son direct et de l'effet de salle.

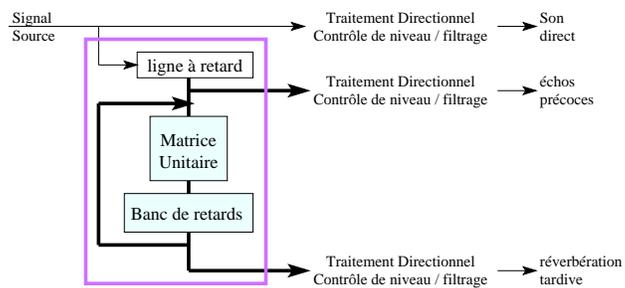


FIG. V.30 – Extension du filtre réverbérant de Stautner et Puckette pour un contrôle des premiers échos préservant le timbre de la réverbération tardive.

Bibliographie

- [1] Y. Ando. *Concert Hall Acoustics*. Springer-Verlag, Berlin, 1985.
- [2] J. Benson. *Audio Engineering Handbook*. mcGraw-Hill, New York, 1988.
- [3] B. A. Blesser and J. M. Kates. Digital processing of audio signals. In Alan V. Oppenheim, editor, *Applications of Digital Signal Processing*, chapter 2. Prentice-Hall, Englewood Cliffs, 1978.
- [4] R. E. Crochiere and L. R. Rabiner. *Multirate digital signal processing*. Prentice-Hall, 1983.
- [5] P. Damaske and Y. Ando. Interaural cross-correlation for multichannel loudspeaker reproduction. *Acustica*, 27:232–238, 1972.
- [6] J. Dattorro. Using digital signal processor chips in a stereo audio time compressor/expander. *Proc. 83rd AES Convention, New York*, Oct 1987. preprint 2500 (M-6).
- [7] G. Fairbanks, W.L. Everitt, and R.P. Jaeger. Method for time or frequency compression-expansion of speech. *IEEE Trans. Audio Electroacoust.*, AU-2:7–12, Jan 1954.
- [8] M. A. Gerzon. Unitary (energy preserving) multichannel networks with feedbacks. *Elec. Letters*, 12(11), 1976.
- [9] A. Gilloire and J.P. Jullien. L'acoustique des salles dans les télécommunications. *L'écho des recherches*, (127):43–54, 1987.
- [10] D. Griesinger. Practical processors and programs for digital reverberation. *Proc. 7th A.E.S Int. Conf.*, pages 187–195, 1989.
- [11] J.M. Jot. *Etude et réalisation d'un spatialisateur de sons par modèles physiques et perceptifs*. PhD thesis, ENST, Sep 1992.
- [12] J. M. Jot. An analysis/synthesis approach to real-time artificial reverberation. *Proc. IEEE ICASSP-92, San Francisco*, Mar 1992.
- [13] H. Kuttruff. *Room acoustics*. Applied Science, London, 1991.
- [14] J. Laroche. Autocorrelation method for high quality time/pitch scaling. *IEEE ASSP Workshop on app. of sig. proc. to audio and acous.*, 1993.
- [15] F. Lee. Time compression and expansion of speech by the sampling method. *J. Audio Eng. Soc.*, 20(9):738–742, 1972.
- [16] G. W. McNally. Dynamic range control of digital audio signals. *J. Audio Eng. Soc.*, 32(5), May 1984.
- [17] J.A. Moorer. About this reverberation business. *Computer Music J.*, 3(2):13–18, 1979.
- [18] M. Puckette. Private communication.
- [19] D. Rossum. Constraint based audio interpolators. *IEEE ASSP Workshop on app. of sig. proc. to audio and acous.*, 1993.
- [20] S. Roucos and A. M. Wilgus. High quality time-scale modification of speech. *Proc. IEEE ICASSP-85, Tampa*, pages 493–496, Apr 1985.

- [21] N. Sakamoto, A. Kurashashi, and Y. Edahiro. A digital audio reverberator. *Proc. 70th A.E.S Conv., N.Y*, 1981. preprint 1810 (I-7).
- [22] M.R. Schroeder and H. Kuttruff. On frequency response curves in rooms. comparison of experimental, theoretical and monte carlo results for the average frequency spacing between maxima. *J. Acoust. Soc. Am.*, 34(76), 1962.
- [23] M.R. Schroeder. Natural sounding artificial reverberation. *J. Audio Eng. Soc.*, 10(3):219–223, 1962.
- [24] M.R. Schroeder. New method for measuring reverberation time. *J. Acoust. Soc. Am.*, 37:409–412, 1965.
- [25] M.R. Schroeder. Digital simulation of sound transmission in reverberant spaces. *J. Acoust. Soc. Am.*, 47(2):424–431, 1970.
- [26] M.R. Schroeder. Progress in architectural acoustics and artificial reverberation: concert hall acoustics and number theory. *J. Acoust. Soc. Am.*, 32:194–203, 1984.
- [27] J. Sikorav. Implementation of reverberators on digital signal processors. *Proc. 80th A.E.S. Conv., Montreux*, 1986. preprint 2326 (C-4).
- [28] J.O. Smith. A new approach to digital reverberation using closed waveguide networks. *Proc. of International Computer Music Conference*, pages 47–63, 1985.
- [29] J. Smith. An allpass approach to digital phasing and flanging. *Proceedings of 1984 International Computer Music Conference, Paris*, 1984.
- [30] J. Stautner and M. Puckette. Designing multi-channel reverberators. *Computer Music J.*, 6(3):52–65, 1982.
- [31] B. Sylvestre and P. Kabal. Time-scale modification of speech using an incremental time-frequency approach with waveform structure compensation. *Proc. IEEE ICASSP-92*, pages 81–84, 1992.
- [32] W. Verhelst and M. Roelands. An overlap-add technique based on waveform similarity (wsola) for high quality time-scale modification of speech. *Proc. IEEE ICASSP-93, Minneapolis*, pages 554–557, Apr 1993.