

## Projet auralisation

---

# Comparaison de différents facteurs sur le procédé d'auralisation

L2 SPI

Année universitaire 2012-2013

# Sommaire

---

<b>I</b>	<b>Etat de l'art</b>	<b>3</b>
<b>II</b>	<b>Principe de l'auralisation</b>	<b>4</b>
II.1.	Produit de convolution . . . . .	4
II.2.	Mise en relation avec la transformée de Fourier . . . . .	5
II.3.	Application à l'auralisation . . . . .	6
<b>III</b>	<b>Essais et Informations Préliminaires</b>	<b>7</b>
III.1.	Réduction du temps de calcul : procédé de convolution . . . . .	7
III.2.	Fenêtrage temporel des réponses impulsionnelles . . . . .	9
III.3.	Type de sons pour les tests . . . . .	11
III.4.	Restitution et écoute post-auralisation . . . . .	11
III.5.	Salles auralisées . . . . .	12
<b>IV</b>	<b>Influence de la source</b>	<b>13</b>
IV.1.	Qualité d'une RI . . . . .	13
IV.1.A.	Amélioration de la qualité de la RI . . . . .	13
IV.2.	Influence du ballon de baudruche . . . . .	14
IV.3.	Compensation de la source . . . . .	15
<b>V</b>	<b>Résultats et interprétations</b>	<b>17</b>
V.1.	Essai 1 : une chanson en salle Mersenne . . . . .	17
V.1.A.	Comparaison perceptive . . . . .	17
V.1.B.	Comparaison fréquentielle . . . . .	17
V.1.C.	Comparaison temporelle . . . . .	18
V.2.	Essai 2 : des claquements de mains en salle réverbérante . . . . .	19
V.2.A.	Comparaison temporelle . . . . .	19
V.2.B.	Comparaison perceptive . . . . .	19
V.3.	Ecoute par d'autres personnes . . . . .	20
<b>A</b>	<b>Spectres pleine échelle en salle Mersenne</b>	<b>22</b>

# Introduction

Pour le public, l'acoustique est un domaine s'appliquant principalement aux salles. Bien que la majorité pense à l'amélioration des performances acoustiques, l'étude réelle autour des salles va beaucoup plus loin. La reproduction des conditions d'écoute dans une salle donnée (réelle ou virtuelle) est un sujet important. Il s'agit d'une application à la frontière entre acoustique des salles et réalité virtuelle, le tout teinté de psychoacoustique. Dans les domaines s'approchant, on citera notamment la reproduction de transducteurs (en captation ou reproduction) et la modélisation de son dits "en 3 dimensions".

Le fait de recréer la modification d'un son par une salle à partir de mesures ou de calculs s'appelle l'*auralisation*. L'auralisation est d'ailleurs définie ainsi dans l'ouvrage de Michael Vorländer [Vor08] :

L'auralisation est une technique visant à créer des fichiers sonores écoutables depuis des données (simulées, mesurées ou synthétisées) numériques.

Afin de mieux saisir le principe et l'importance des différents facteurs en jeu dans le processus d'auralisation, un bref historique est dressé puis une étude théorique des outils mathématiques est menée ; celle ci conduit à la première auralisation pratique présentée. Divers essais préliminaires sont consignés ensuite avant de discuter l'influence que la source a sur le procédé, une tentative de compensation est aussi décrite. Finalement, les résultats de deux auralisations sont discutés ; cette dernière partie accueillant aussi diverses remarques sur la différence entre réponses impulsionnelles – RI – binaurales (deux points de captation de part et d'autre d'une tête artificielle) et monaurales (un unique micro pour la captation).

## Sources et données du projet

---

Toutes les sources du projet (,  $\text{\LaTeX}$ , etc...) ainsi que les données de mesures et les résultats sont disponibles en ligne :

[https://github.com/Matael/projet\\_auralisation\\_12](https://github.com/Matael/projet_auralisation_12)

# Chapitre I

## Etat de l'art

Les premiers essais se rapportant à l'auralisation ont été faits par Spandöck et coll. en 1929. Ces travaux ont été repris et améliorés après l'apparition des ordinateurs ; en utilisant cette nouvelle puissance de calcul et, vers la fin des années 1960, le premier logiciel de simulation d'acoustique des salles fut développé (Krokstad) [Vor08].

Le mot «auralisation» lui-même fut utilisé pour la première fois par Kleiner et coll. dans l'article Auralization – An Overview [KDS93].

Dans la somme des techniques utilisées pour aboutir à la reproduction des conditions acoustiques d'une salle, deux reviennent principalement :

- utilisation de *ray-tracing* ;
- utilisation de systèmes source-image.

Les deux sont connues depuis longtemps et éprouvées, elles peuvent parallèlement être améliorées *via* la prise en compte de divers facteurs :

- diffusion (aléatoire ou déterminée) ;
- absorption.

Les applications possibles de l'auralisation au terme général sont nombreuses et variées. La plus évidente d'entre elles est certainement la reproduction de «l'acoustique» d'une salle, mais on peut aller plus loin. L'acoustique prédictive permet de simuler le rendu de salles et plus généralement d'espaces inexistantes en combinant des mesures entre elles ou par pure simulation. Enfin, l'auralisation peut être utilisée dans la conception de systèmes de réalité virtuelle à forte immersion.

Il faut enfin savoir que la complexité de restitution d'une auralisation rend difficile sa mise en place à grande échelle. Si la projection en 3D et à 360 degrés est aujourd'hui possible *via* divers processus de visualisation (le pendant visuel de l'auralisation), l'inclusion d'un environnement acoustique pleinement contrôlé est ardue et demande un nombre impressionnant de haut-parleurs (et ne pourrait s'adapter à chaque spectateur). Une restitution du champ acoustique de manière individuelle peut être en revanche envisageable au travers de casques et d'un système de suivi d'orientation 3D. Ce type d'environnement d'immersion existe soit en version prototypée (en Allemagne par exemple, à Ilmenau) ou bien en version commerciale *via* le projet CAVE<sup>TM</sup><sup>1</sup>.

---

1. CAVE<sup>TM</sup> : CAVE<sup>TM</sup> Automatic Virtual Environment

## Chapitre II

# Principe de l'auralisation

Le but de cette section est de mettre en lumière le procédé mathématique permettant l'auralisation d'une salle. Pour cela nous allons nous intéresser à plusieurs notions qui rentrent en jeu dans l'auralisation.

### II.1. Produit de convolution

---

Ce projet traite uniquement de salles à taille réelle avec une source et un récepteur fixes. Dans ce cas particulier, la salle à étudier peut être assimilée à un filtre linéaire invariant par translation dans le temps.

Pour qu'un système puisse être considéré comme linéaire, il suffit que pour des entrées  $x_1(t)$  et  $x_2(t)$  et leurs sorties respectives  $y_1(t)$  et  $y_2(t)$  on ait :

$$\alpha x_1(t) + \gamma x_2(t) \Leftrightarrow \alpha y_1(t) + \gamma y_2(t) \quad (\text{II.1})$$

Dans notre cas, si 2 sons d'enveloppes et d'amplitudes différentes sont émis dans une salle, il paraît logique que ces sons n'interagissent pas entre eux et que par conséquent cette équation soit vérifiée dans le cas de l'émission d'un son dans une salle avec un émetteur et un récepteur fixes.

On peut dire qu'un système est invariant par translation dans le temps si, alors qu'à un temps  $t$  une entrée  $x(t)$  est reliée à une sortie  $y(t)$  on a pour un temps  $t + \tau$  une entrée  $x(t + \tau)$  liée à une sortie  $y(t + \tau)$ .

Dans le cadre de ce rapport, à la condition que ni l'émetteur ni le récepteur ne changent de position et que les conditions extérieures ne fluctuent pas trop (température, pression ambiante), la réponse d'une salle à un son donné n'a *a priori* aucune raison de varier dans le temps. Une salle peut donc bien être approximée à un système linéaire. On a donc le système présenté figure II.1.

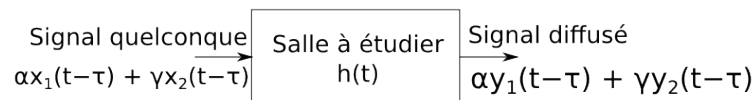


FIGURE II.1 – Le système à étudier est considéré invariant par translation dans le temps si on peut lier les entrées aux sorties par une fonction de transfert constante pendant le temps considéré.

Comme les systèmes étudiés sont tous linéaires, il semble intéressant de rappeler certaines des propriétés de ces systèmes. Sachant que le système est linéaire, on a donc :

$$\alpha x_1(t) + \gamma x_2(t) \Leftrightarrow \alpha y_1(t) + \gamma y_2(t) \quad (\text{II.2})$$

Et

$$x_1(t) \Leftrightarrow y_1(t)x_1(t - \tau) \Leftrightarrow y_1(t - \tau) \quad (\text{II.3})$$

En combinant ces 2 propriétés on peut en déduire que :

$$\alpha x_1(t - \tau) + \gamma x_2(t - \tau) \Leftrightarrow \alpha y_1(t - \tau) + \gamma y_2(t - \tau) \quad (\text{II.4})$$

On considère maintenant un signal quelconque  $e(t)$ . On peut approximer ce signal par une somme de signaux impulsionnels  $a_i(t)$  d'amplitudes différentes et décalés dans le temps (peigne de Dirac). On a donc :

$$e(t) = \sum_i A_i a(t - \tau_i) \quad (\text{II.5})$$

Comme le système est linéaire on peut donc en déduire que la sortie du système s'écrira :

$$s(t) = \sum_i A_i h(t - \tau_i) \quad (\text{II.6})$$

Si on passe cette écriture à la limite continue [Mar12], on obtient :

$$e(t) \rightarrow s(t) = \int e(\tau) h(t - \tau) d\tau \quad (\text{II.7})$$

Cette dernière relation est fondamentale et est nommée «produit de convolution». Ce produit est dénoté par le signe  $*$ . De plus on constate que la fonction  $h(t)$  correspond à la sortie du système pour une unique impulsion envoyée en entrée (visible dans le cas où on prend un  $i$  unique). Cette fonction est essentielle dans le processus d'auralisation, il s'agit de la fonction de transfert caractérisant le système dans le domaine temporel, aussi nommée réponse impulsionnelle du système (RI).

## II.2. Mise en relation avec la transformée de Fourier

---

En termes de ressources et de temps de calcul, le produit de convolution est extrêmement lourd ; il est, de plus, peu maniable. Il est toutefois possible de se servir d'une autre opération mathématique afin de rendre plus simple l'utilisation du produit de convolution. La transformée de Fourier et l'espace de Fourier proposent des propriétés intéressantes. Des algorithmes bien connus permettent de plus d'alléger le calcul et de l'accélérer (*Fast Fourier Transform* notamment). La transformée de Fourier est définie de la manière suivante :

$$\mathcal{F}\{x(t)\} = \int_{-\infty}^{+\infty} x(t) e^{-2i\pi Ft} dt \quad (\text{II.8})$$

Une des propriétés intéressantes concerne la transformée de Fourier d'un produit de convolution :

$$\begin{aligned}
\mathcal{F}\{x(t) * y(t)\} &= \int_{-\infty}^{+\infty} [x(t) * y(t)] e^{-2i\pi Ft} dt \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x(u)y(t - \tau) du e^{-2i\pi Ft} dt \\
&= \int_{-\infty}^{+\infty} \left[ \int_{-\infty}^{+\infty} x(u)y(t - \tau) e^{-2i\pi Ft} \right] du dt
\end{aligned}$$

On pose  $v = t - u$ , on a donc  $t = u + v$  à  $u$  fixé,

$$\begin{aligned}
\mathcal{F}\{x(t) * y(t)\} &= \int_{-\infty}^{+\infty} \left[ \int_{-\infty}^{+\infty} x(u)y(v) e^{-2i\pi F(v+u)} \right] du dv \\
&= \int_{-\infty}^{+\infty} x(u) e^{-2i\pi Fu} du \int_{-\infty}^{+\infty} y(v) e^{-2i\pi Fv} dv \\
&= \hat{x}(F) \cdot \hat{y}(F)
\end{aligned}$$

La transformée de Fourier du produit de convolution de 2 signaux est donc égale à la multiplication des transformées de Fourier de chacun des signaux. Par conséquent, les calculs de produits de convolutions sont faits en passant par le domaine de Fourier pour l'accélération des calculs (voir le paragraphe III.1.).

## II.3. Application à l'auralisation

---

Dans le cas d'un système linéaire on a :

$$e(t) * h(t) = s(t) \quad (\text{II.9})$$

Avec  $e(t)$  le signal en entrée du système,  $h(t)$  sa RI et  $s(t)$  le signal en sortie du système pour  $e(t)$  en entrée. Dans le cas de l'auralisation d'une salle, le signal d'entrée correspond au signal anéchoïque à émettre dans la salle et le signal de sortie au signal enregistré une fois le système excité. Lorsque la réponse impulsionnelle est connue (celle-ci étant facilement mesurable), il est possible, à partir de celle-ci et d'un son anéchoïque, d'en déduire le son tel qu'il pourrait être perçu si le signal anéchoïque avait été réellement émis dans la salle. C'est cette opération qu'on nomme auralisation.

Il faut cependant prendre en compte d'autres phénomènes dans le procédé d'auralisation, qui seront dûs au fait que notre source impulsionnelle excitatrice et la chaîne de mesure ne soient pas parfaites. C'est lors de l'application des compensations pour les sources et chaîne de mesure que passer dans le domaine de Fourier sera réellement utile pour limiter la complexité des calculs de convolution par simples multiplications et divisions de spectres.

## Chapitre III

# Essais et Informations Preliminaires

Avant de mettre en place des auralisations concrètes, une étude des moyens de comparaison et divers essais préliminaires sont menés.

### III.1. Réduction du temps de calcul : procédé de convolution

---

La convolution est une opération mathématique très gourmande en temps processeur et en mémoire. Même si elle peut être assez facilement parallélisée<sup>1</sup>, ces deux solutions étaient hors de portée ici.

La convolution joue dans le projet un rôle central puisqu'elle est l'outil permettant de passer d'une RI et d'un son anéchoïque à un résultat sonore représentant la façon dont l'espace lié à la RI aurait modifié le son (voir figure III.1).

La fonction `conv()` proposée par MATLAB® et GNU/Octave réalise une convolution mathématique stricte (au sens discret). Celle-ci est toutefois très lente sur des machines peu puissantes.

La fonction `fftconv()` proposée par GNU/Octave est par contre nettement plus rapide, elle utilise le mécanisme présenté en bas de la figure III.1. Cette fonction n'est toutefois pas proposée nativement dans MATLAB®. Elle est réimplémentée pour le projet en s'appuyant sur un script trouvé sur le site utilisateurs du logiciel<sup>2</sup> :

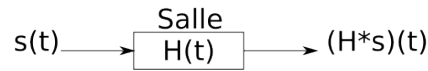
---

1. Soit avec des processeurs mathématiques dédiés soit avec des processeurs hautement parallèles type GPU (processeurs de carte graphique)

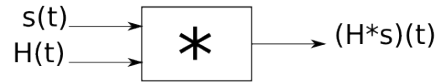
2. <http://www.mathworks.com/matlabcentral/fileexchange/5703-fftconv/content/fftconv.m>



### Convolution "physique"



### Convolution mathématique 1



### Convolution mathématique 2

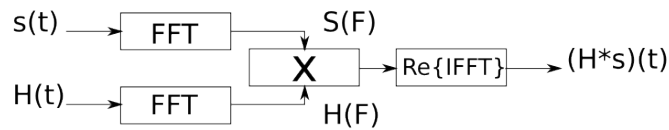


FIGURE III.1 – Lien entre le projet et la convolution. En haut, la convolution «physique» (lorsque qu'un signal  $s(t)$  est émis dans la salle (de réponse impulsionnelle  $H(t)$ ). Au milieu la convolution mathématique au sens strict (*via* l'opération dédiée). En bas, la convolution mathématique en utilisant l'équivalence entre convolution temporelle et multiplication fréquentielle.

```
function c = fftconv(a,b);
```

```
na = length(a);
```

```
nb = length(b);
```

```
n=na+nb;
```

```
A = fft(a,n);
```

```
B = fft(b,n);
```

```
c = ifft(A.*B,n);
```

```
c = real(c(1:na+nb-1));
```

La transformée de Fourier (TF) est une opération qui n'est pas strictement réversible et qui entraîne une légère perte de données :

$$\mathcal{F}^{-1} \{ \mathcal{F} \{ s(t) \} \} \approx s(t)$$

Il semble intéressant de s'intéresser (au moins un peu) à l'erreur induite par l'utilisation de `fftconv()` plutôt que `conv()`. La même convolution est réalisée avec chacune des fonctions, puis une calcul de distance géométrique est utilisée pour déterminer la différence existant entre les deux signaux :

$$d(t) = \sqrt{|[s_1(t)]^2 - [s_2(t)]^2|}$$

où  $d(t)$  est la distance entre les signaux  $s_1$  et  $s_2$  au point d'abscisse  $t$ .

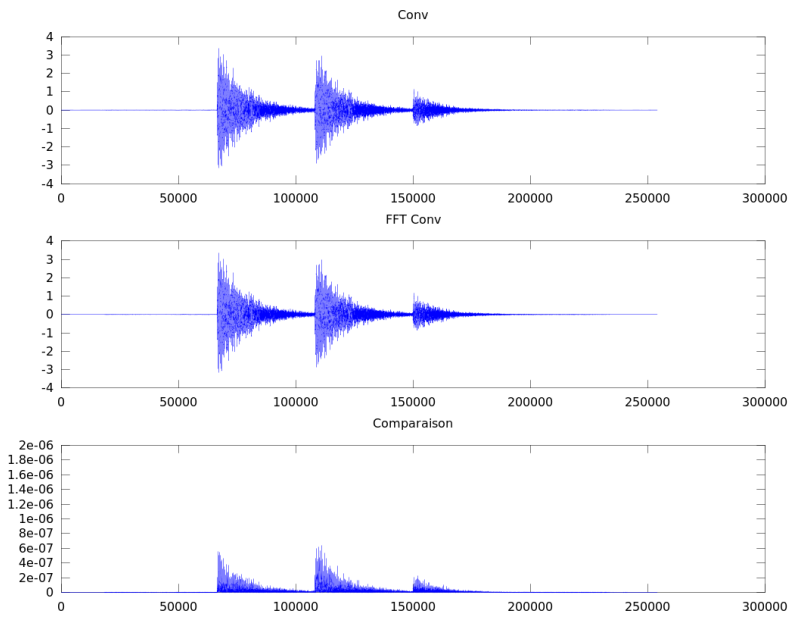


FIGURE III.2 – Comparaison entre les résultats d’une même convolution avec `conv()` d’une part et `fftconv()` d’autre part. On remarque notamment que sur le graphe du bas (représentant la distance géométrique entre les 2 graphes au dessus), l’axe des ordonnées est gradué entre  $2 \cdot 10^{-7}$  et  $2 \cdot 10^{-6}$ .

Le graphe obtenu est présenté en figure III.2, il faut notamment noter que l’axe des ordonnées sur le graphe des distances est gradué entre  $2 \cdot 10^{-7}$  et  $2 \cdot 10^{-6}$ .

Devant la faible amplitude de l’erreur provoquée par l’utilisation de `fftconv()` et le gain en terme de temps de calcul, cette fonction semble largement plus avantageuse que `conv()`.

## III.2. Fenêtrage temporel des réponses impulsionnelles

Lors de la prise de réponses impulsionnelles, afin de ne pas dégrader l’information en «coupant» les réflexions les plus tardives, les mesures sont faites sur des temps assez longs (voir figure III.3).

De telles mesures posent plusieurs soucis, d’abord en terme de stockage mais aussi en terme de temps de calcul.

Les mesures de RI sont donc fenêtrées pour éliminer les blancs avant et après le signal utile. Afin de conserver les fichiers originaux et d’éviter l’apparition d’incohérences dans les fichiers de mesures, les fenêtrages sont paramétrés à la main et appliqués dans les scripts de traitement et les fichiers de mesures sont laissés tels quels.

Il semble intéressant enfin de regarder quelle influence a le fenêtrage d’une RI sur le processus d’auralisation.

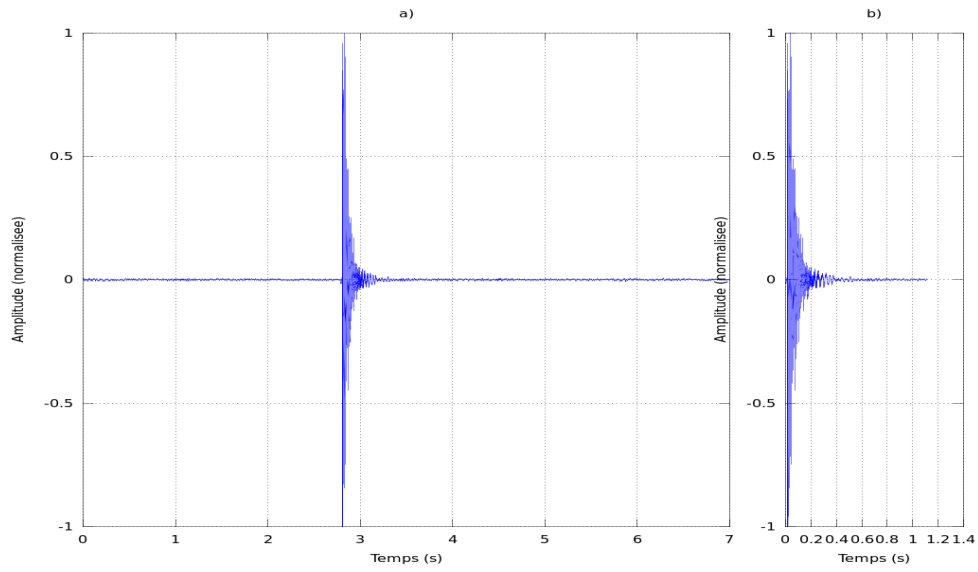


FIGURE III.3 – Une des RI mesurées avant le fenêtrage (ici, le canal gauche d’une RI binaurale en salle Mersenne). La figure a) montre la RI avant fenêtrage et la figure b) la même RI après fenêtrage

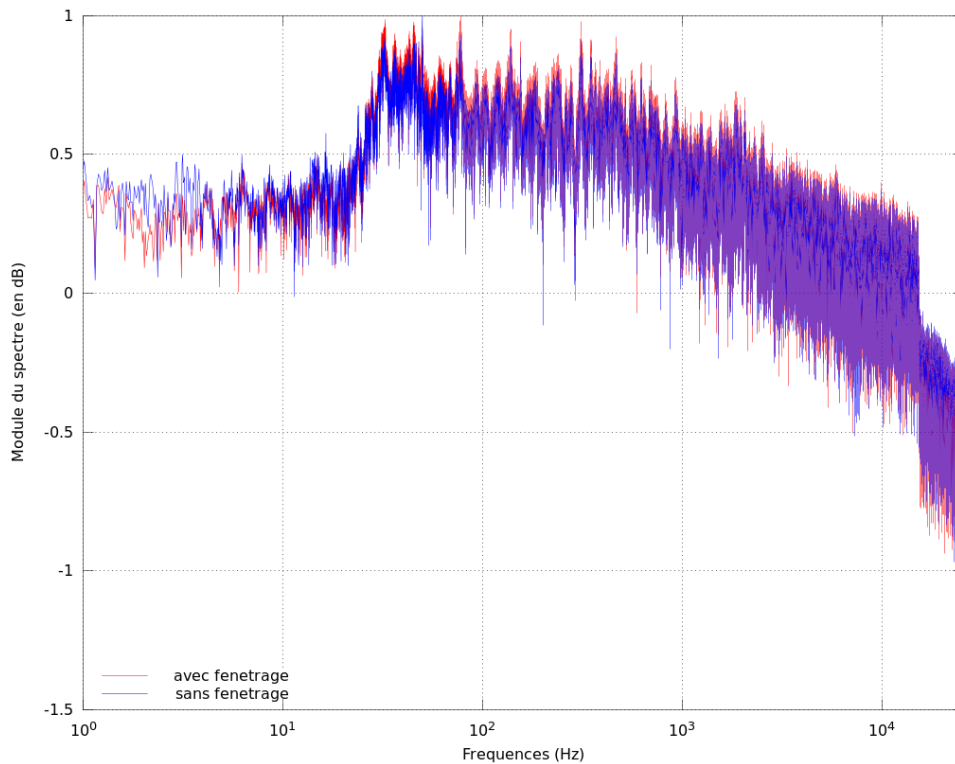


FIGURE III.4 – Les spectres des sons resultant de la convolution avec une RI fenêtrée (en rouge) et non fenêtrée (en bleu). Même si les écarts sont faibles, ils existent.

Après deux tests, il apparaît que le fenêtrage des RI apporte un gain non négligeable en terme de temps de calculs<sup>3</sup>. Toutefois, comme le montre la figure III.4, on remarque que les spectres de signaux convolués avec une RI fenêtrée d’une part et non fenêtrée de l’autre ne sont pas identiques. Il faut malgré tout préciser que perceptivement l’utilisation d’une RI fenêtrée rend le son plus net et moins bruité (plus réaliste en fait). Cela vient probablement de l’élimination par fenêtrage du bruit de fond avant et après le son utile.

### III.3. Type de sons pour les tests

---

Le choix des sons de test est un choix assez important dans le sens où il faut que ceux-ci soient suffisamment génériques pour ne pas biaiser les résultats mais assez particuliers pour que les altérations produites par la salle cible soient audibles (ou visibles).

D’après Kleiner et coll., la parole ou la musique sont de bons sons de test [KDS93]. le choix des sons s’est donc porté sur :

- une suite de claquements de mains (pour la composante sourde et impulsionnelle) ;
- des tintements de clés ;
- les 30 premières secondes du morceau *Mon Imagination* de Pierpoljak.

Les deux premiers sons proviennent du site de partage de sons *Freesound*<sup>4</sup> ; ils ont été enregistrés en salle anéchoïque et échantillonnés à 96 kHz. Ils sont ré-échantillonné à 51200Hz (la fréquence utilisé par *Analyseur CTTM* pour les mesures).

La chanson quant à elle provient d’un album studio et l’enregistrement est donc teinté de la signature de la salle d’enregistrement.

### III.4. Restitution et écoute post-auralisation

---

Le succès d’une auralisation repose grandement sur les conditions de l’écoute finale. En effet, il faut que le son à écouter soit le moins altéré possible par l’environnement d’écoute. Deux moyens d’écoute existent (notamment dans le cas d’écoute de résultats binauraux) : au casque ou *via* des haut-parleurs et un système de restitution.

Si la première méthode ne pose pour ainsi dire aucun problème, la seconde est nettement plus complexe. Comme il s’agit d’une auralisation binaurale, les deux canaux (droite et gauche) sont différenciés et il est important que chaque oreille ne capte que ce qui lui est destiné.

Il y a donc des règles à respecter au niveau du système de restitution avec notamment un système anti-diaphonie [KDS93]. De tels systèmes sont parfois difficiles à mettre en place notamment dans des salles assez grandes, ceci étant dû au second problème soulevé par une restitution hors casque : il faut que l’empreinte acoustique soit faible pour ne pas perturber le son émis [Bru10].

---

3. une partie des calculs étant faits sur une machine peu puissante, cette composante est importante

4. <http://www.freesound.org/>, merci à Anton (<http://www.freesound.org/people/Anton/>) d’avoir posté ces sons là.

### III.5. Salles auralisées

---

Deux salles sont utilisées au cours du projet :

- la salle de TP Mersenne (voir figure III.5) ;
- la salle réverbérante (à côté de la salle mersenne).

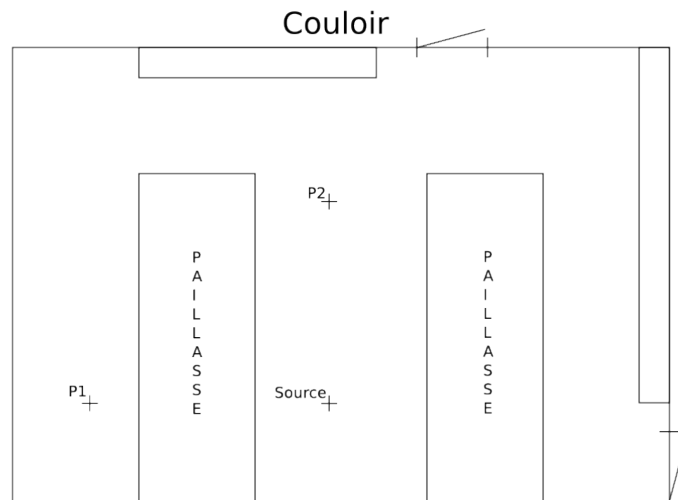


FIGURE III.5 – Plan de la salle Mersenne. La source est été placée au point noté **Source**. Le micro (pour les prises monaurales) et la tête sont toujours placés en **P1** ou **P2**. L'orientation par défaut de la tête est vers le couloir pour **P1** et vers la source pour **P2**. La source est toujours orientée vers le récepteur.

# Chapitre IV

## Influence de la source

Dans le processus d'auralisation, la prise de réponse impulsionnelle ou le calcul de celle-ci est un passage important.

En réalité, la qualité du signal final dépend grandement de la qualité de la RI.

### IV.1. Qualité d'une RI

---

L'estimation de la qualité d'une RI passe notamment par 3 facteurs :

- la netteté de l'attaque ;
- un bon rapport signal/bruit (RSB) ;
- la qualité de l'impulsion.

La netteté de l'attaque et la qualité de l'impulsion sont liées en fait, mais une impulsion bien générée sera toujours tributaire de la qualité de la source, en cela, cette dernière est clairement le facteur limitant.

Un bon RSB est parfois assez difficile à obtenir (milieu bruyant, parasites, etc...), il peut toutefois être amélioré.

#### IV.1.A. Amélioration de la qualité de la RI

Plusieurs techniques peuvent permettre d'améliorer les différents paramètres d'une RI. On peut d'ailleurs regrouper ces techniques en deux ensembles : celles agissant sur le procédé de mesure lui-même et celles agissant *a posteriori*.

La seconde classe comporte de nombreuses techniques de traitement du signal et de réduction du bruit :

- déconvolution de signaux ;
- filtrage simple ;
- filtrage Wiener.

Pour ce qui est des techniques interférant dans le procédé de mesure, on peut notamment citer la prise de réponses fréquentielles plutôt qu'impulsionnelles ou l'utilisation de signaux MLS et de systèmes spécialisés (système CLIO<sup>TM</sup> par exemple).

La prise de réponses fréquentielles permet d'utiliser des enceintes avec de multiples hauts parleurs, même si ceux ci n'ont pas des temps de réponse comparables. En effet, il s'agit d'émettre un bruit le plus large bande possible et de calculer la fonction de transfert dans le domaine fréquentiel :

$$H(F) = \frac{S(F)}{E(F)}$$

où  $H(F)$  est la fonction de transfert en fréquentiel,  $E(F)$  le signal excitateur et  $S(F)$  le signal mesuré dans la salle.

On peut ensuite repasser dans le domaine temporel pour récupérer une réponse impulsionnelle en appliquant une transformée de Fourier inverse sur  $H(F)$ .

Les deux avantages de l'utilisation de réponses fréquentielles c'est qu'elles sont moyennables (ce qui permet de réduire les erreurs) et qu'il est possible de les mesurer à fort niveau, améliorant ainsi le RSB. Il faut toutefois faire attention à ce que toutes les fréquences soient excitées de manière semblable (là encore, les transducteurs sont limitants).

La deuxième possibilité, impliquant des systèmes spécialisés, permet de mesurer des RI de bonne qualité même dans des environnements (relativement) bruyants. C'est un avantage considérable de l'utilisation de signaux MLS (*Maximum Length Sequence*). Le système CLIO<sup>TM</sup> permet ce genre de choses mais comme tout système spécialisé, il demande un certain temps de prise en main. Il n'a pas été utilisé au cours de ce projet.

## IV.2. Influence du ballon de baudruche

---

La réponse impulsionnelle d'une salle doit la caractériser et particulièrement les différentes réflexions induites par sa géométrie.

Afin d'avoir une mesure la plus fidèle possible à la réalité du terrain, il faut que la source utilisée pour générer l'impulsion (ou le bruit permettant la prise d'une réponse en fréquences) soit la plus omni-directionnelle possible.

L'utilisation d'un ballon de baudruche posait le souci que sa directivité est inconnue. Avec plus de temps, une caractérisation de la directivité du ballon aurait pu être intéressante.

Une telle étude semble être en cours ou avoir été réalisée au LIMSI [Bru10], malheureusement l'étude n'est pas disponible à la consultation : pour ce projet, le ballon de baudruche sera donc considéré omnidirectionnel.

Le second inconvénient soulevé par l'utilisation d'un ballon de baudruche est que la bande passante du son généré par son éclatement n'est pas connue. Un éclatement de ballon de baudruche est donc mesuré en salle semi-anéchoïque (le spectre est visible en figure IV.1).

La réponse en fréquence n'est donc clairement pas plate. On note une décroissance de 20dB par décade et une série de minima en hautes fréquences ainsi qu'un creux dans la bande 10 - 200Hz.

Pour mieux saisir la correspondance entre la bande passante du ballon et les résultats à l'écoute, on observe le graphe de la valeur absolue au carré du spectre. Cela nous donne le spectre d'énergie présenté en figure IV.2.

On remarque que l'énergie se situe principalement dans les basses fréquences (en deçà de 100Hz) et que 2 pics viennent potentiellement perturber la RI entre 200 et 400Hz.

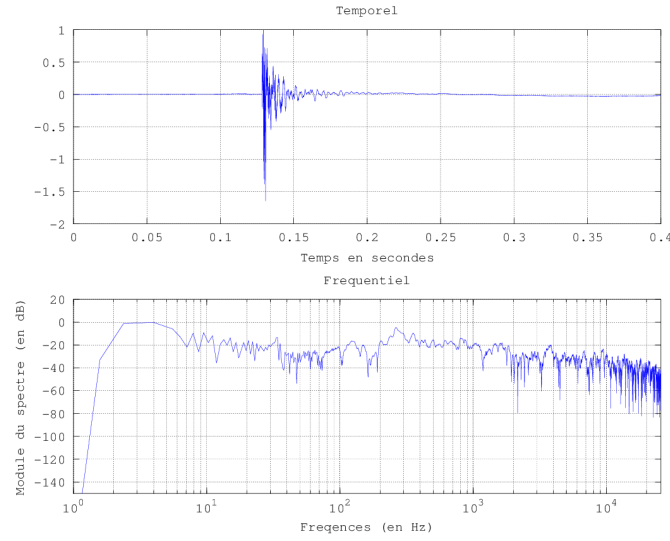


FIGURE IV.1 – Enregistrement temporel de l'éclatement d'un ballon de baudruche en salle semi-anéchoïque (noter les réflexions visibles en temporel) et spectre en dB du signal.

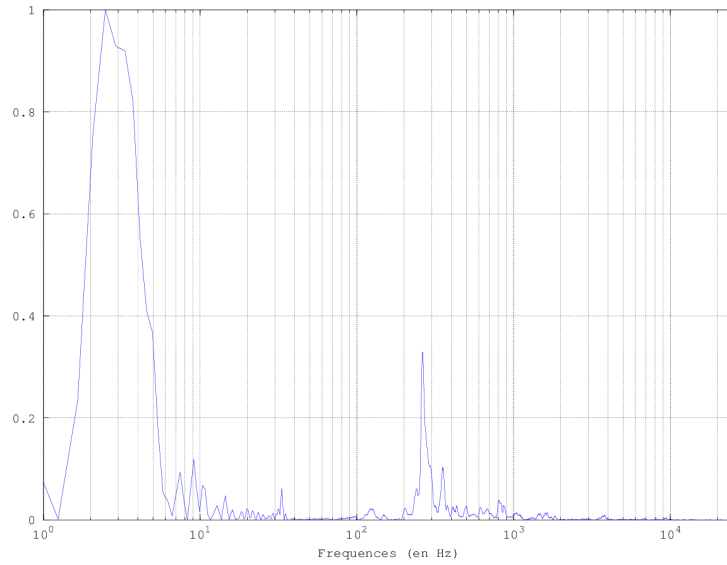


FIGURE IV.2 – Spectre d'énergie d'un ballon de baudruche en salle semi-anéchoïque.

### IV.3. Compensation de la source

Les sources utilisées ne sont donc pas parfaites, cependant il est possible de les prendre en considération lors du calcul du son convolué. En effet, les imperfections dues à la chaîne d'excitation peuvent être modélisées par un filtre qui viendrait «avant» la salle modélisée par un système linéaire. De plus la chaîne d'acquisition est elle supposée parfaite; en pratique ce n'est pas tout à fait le cas, mais devant l'imperfection de la chaîne d'excitation l'approximation est correcte. Cette configuration est représentée sur le schéma figure IV.3. Afin de pouvoir compenser la chaîne d'excitation; il faut donc déterminer  $y(t)$ . Cette grandeur correspond, dans le cas du ballon, à la fonction permettant de passer d'une impulsion parfaite à celle donnée par le ballon : il s'agit de la fonction de transfert du ballon. En définitif, pour connaître cette fonction de transfert, une mesure de l'éclatement



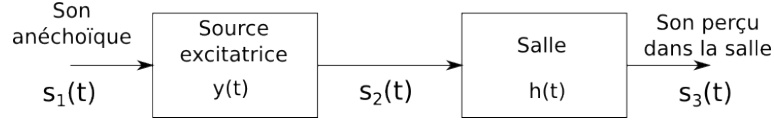


FIGURE IV.3 – Schéma du système étudié avec prise en compte de la chaîne d'excitation

d'un ballon en salle anéchoïque (là où la salle a le moins d'influence) est réalisée. Cette mesure est présentée en figure IV.1. Une fois  $y(t)$  déterminé, il ne reste qu'à en déduire  $h(t)$ . On sait que :

$$\begin{aligned} s_1(t) * y(t) &= s_2(t) \Leftrightarrow S_1(F) \times Y(F) = S_2(F) \\ s_2(t) * h(t) &= s_3(t) \Leftrightarrow S_2(F) \times H(F) = S_3(F) \end{aligned}$$

On peut donc en déduire que :

$$h(t) = \mathcal{F}^{-1} \left\{ \frac{S_3(F)}{S_2(F)} \right\}$$

Bien que retrouver la réponse impulsionnelle de la salle exempte de l'influence du ballon semble simple sur un plan théorique, d'un point de vue pratique, de grosses difficultés ont été rencontrées. En effet, voici en figure IV.4 la réponse impulsionnelle obtenue dans le domaine temporel et fréquentiel :

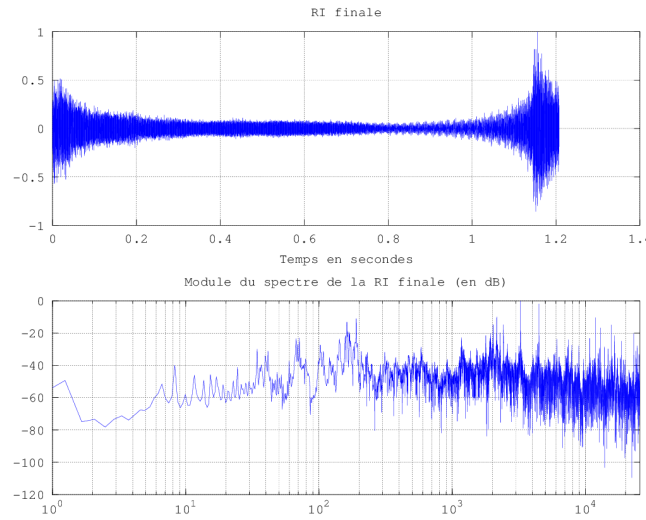


FIGURE IV.4 – Réponse impulsionnelle calculée représentée temporellement et fréquentiellement.

Le résultat ne semble donc pas concluant. Il est probable que le problème vienne des modifications éventuelles qu'apportent la transformée de Fourier au niveau de la phase des signaux. Cependant nos connaissances sur le sujet sont trop limitées pour permettre de se pencher de manière approfondie sur la question.

## Chapitre V

# Résultats et interprétations

La perception sonore humaine est dite binaurale : c'est à dire qu'il y a deux «capteurs» (en l'occurrence de chaque côté de la tête). Cette particularité est importante dans la perception de l'espace, en effet, le volume de la tête retarde la propagation du son tout en déformant celui-ci permettant ainsi un repérage dans le plan horizontal (avec une précision pouvant aller jusqu'à un degré [Vor08]). Le torse a lui aussi une influence, particulièrement pour le repérage dans le plan vertical. Au cours des mesures pour ce projet, une tête artificielle (sans torse) a été utilisée, le repérage vertical sera donc difficile à reproduire d'après nos mesures.

### V.1. Essai 1 : une chanson en salle Mersenne

---

Le premier essai approfondi est réalisé en salle Mersenne. La source est au point **Source** (voir figure III.5) et le récepteur au point **P1** (la tête est dans l'orientation par défaut).

#### V.1.A. Comparaison perceptive

A l'écoute, la différence entre les deux résultats (monaural et binaural) est flagrante. Alors que la position de la source est strictement indéterminable en monaural, elle est bien identifiable en binaural.

Des mesures plus précises auraient probablement permis une meilleure reconnaissance de la géométrie de la salle et des obstacles, en particulier la paillasse présente sur la droite du récepteur.

#### V.1.B. Comparaison fréquentielle

On note par ailleurs la différence de contenu fréquentiel (et en particulier la différence de niveau) sur la figure V.1. Cette différence observée sur les spectres est bel et bien en accord avec la comparaison perceptive menée ci-avant. On remarque que le signal monaural vient effectivement «s'intercaler» entre les 2 canaux du signal binaural (en haut monaural et canal droit, en bas monaural et canal gauche). Une représentation pleine échelle est disponible en annexe.

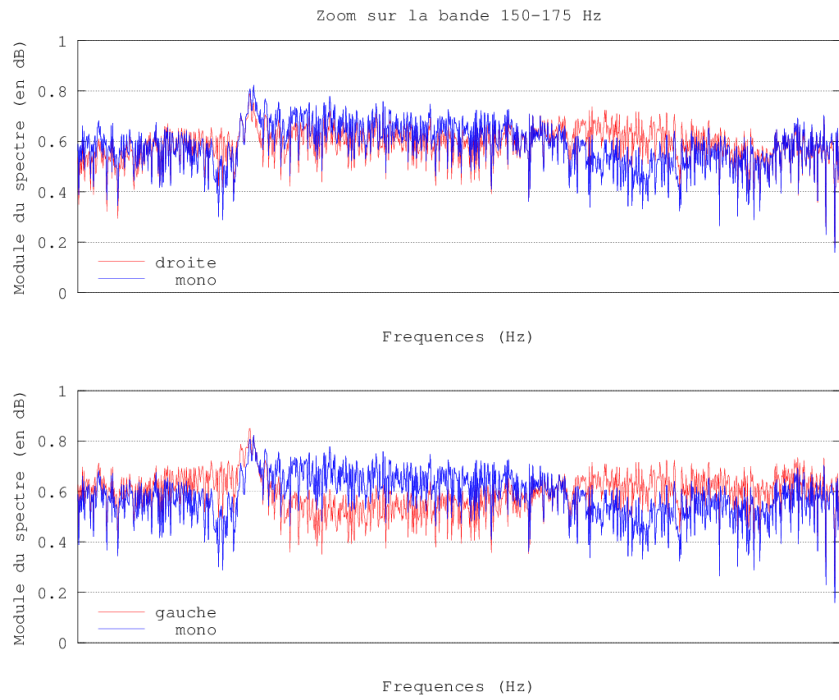


FIGURE V.1 – Comparaison entre les contenus fréquentiels des signaux monaural et binaural dans la bande 150-175Hz.

### V.1.C. Comparaison temporelle

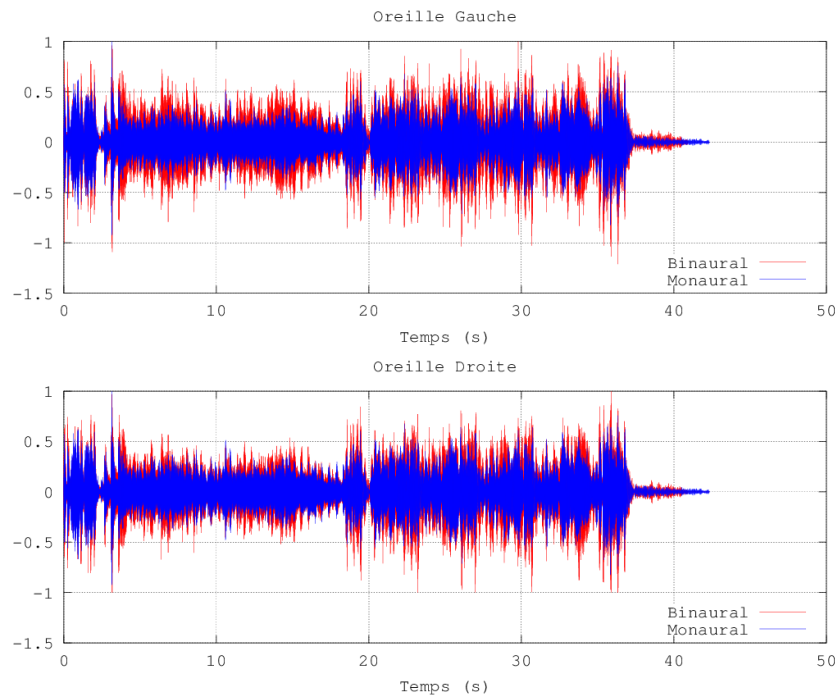


FIGURE V.2 –  
Tracés temporels des signaux convolués pour une RI monaurale (en bleu) et une RI binaurale en rouge. En haut pour l'oreille gauche et en bas pour l'oreille droite.

La figure V.2 montre les tracés temporels des signaux convolués en monaural et binaural. Les différences sont assez minimales (mise à part l'amplitude). Les enveloppes sont très ressemblantes. Un même son est convolué à chaque fois, le résultat est donc censé être ressemblant.

## V.2. Essai 2 : des claquements de mains en salle réverbérante

Un autre essai est mené en salle réverbérante. Il s'agit cette fois de vérifier que les différentes réflexions sont correctement restituées lors de la convolution.

### V.2.A. Comparaison temporelle

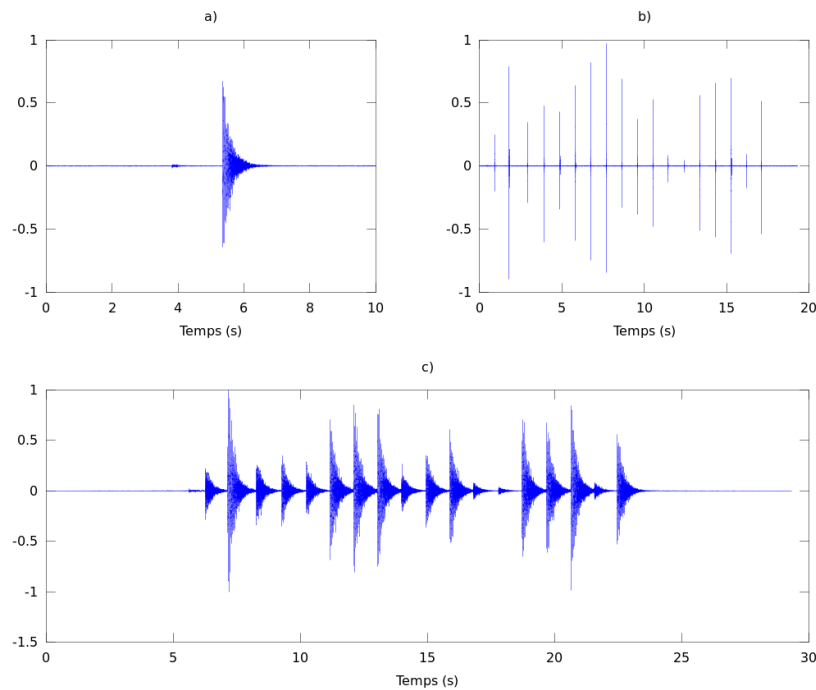


FIGURE V.3 – Tracé des différents signaux et résultat de la convolution en temporel. En a) la RI utilisée, en b) le son anéchoïque, en c) le résultat après convolution des deux signaux précédents.

La figure V.3 montre les différents signaux utilisés et leur convolution. Ici, le résultat est donné en monaural, on note que les réflexions visibles sur la RI sont clairement restituées sur le signal final.

En binaural, le résultat de chaque canal ressemble très fortement au résultat en monoral présenté ci-avant.

### V.2.B. Comparaison perceptive

L'écoute se fait sur les convolutions binaurales.

Perceptivement, le résultat est probant. L'influence de la salle réverbérante s'entend très bien et l'immersion se fait sans souci.

La provenance du son est moins clairement identifiable que pour une salle plus mate (salle Mersenne par exemple), mais toujours nettement plus identifiable qu'en monaural.

### V.3. Ecoute par d'autres personnes

---

Tout au long du projet, les sons résultants ont été écoutés par plusieurs personnes étrangères au projet.

Il ressort plusieurs choses de ces écoutes :

- tous trouvent le résultat étonnant ;
- la plupart (presque tous) réussit à retrouver la provenance du son lors d'écoutes de résultats binauraux ;
- les écoutes monaurales sont troublantes pour ceux les écoutent (incapacité à localiser la source, symétrie auditive inhabituelle) ;
- ceux connaissant la différence entre salle mate et salle réverbérante lient clairement les écoutes aux types de salles, les autres parlent de «*plus de reverb*» pour les résultats utilisant une salle réverbérante ;
- les résultats liés à la salle Mersenne donnent à certains l'impression que ce qu'ils entendent provient de la salle où ils se trouvent et non du casque.

Finalement, la qualité des écouteurs utilisés influe grandement sur le résultat. L'immersion est assez faible pour de simple écouteurs, elle demande une certaine concentration. En utilisant des écouteurs intra-auriculaires, l'immersion est largement meilleure et la provenance est plus facilement identifiable. Les sons semblent plus naturels. Les meilleurs résultats sont obtenus en utilisant un casque isolant convenablement l'auditeur des bruits qui l'entourent. L'immersion est alors bonne.

# Conclusion

La mise en place d'une chaîne d'auralisation (acquisition, convolution et restitution au casque) est un procédé relativement simple et permet des résultats honorables. Les problèmes se présentent lorsqu'une recherche d'optimisation est incluse. Si l'utilisation de RI binaurales améliore grandement la qualité de l'auralisation (en permettant notamment un meilleur repérage spatial), ce n'est pas le seul levier vers un rendu optimisé. L'utilisation d'autres sources qu'une simple boudruches ainsi que d'autres procédés de mesure de RI sont autant de moyen d'aboutir à une auralisation de meilleure qualité. L'étude des salles elles-mêmes n'a pas été traité dans ce rapport mais pourrait présenter un intérêt pour choisir les sources en fonction des salles ciblées, de même une optimisation au regard du type de sons à utiliser ensuite permettrait de réduire les temps de calculs.

L'utilisation de procédés tels que celui-ci dans des applications de réalité virtuelle ou la qualité de l'immersion se doit d'être la plus précise possible est aussi un vecteur de développement. Au cours du projet et notamment de l'analyse de résultats, nous avons noté combien les ressources (processeur, mémoire) nécessaire pouvait croître rapidement. De ce point de vue, l'écriture et la conception de systèmes dédiés à cette problématique semblent intéressantes.

Nos résultats, s'ils sont loin d'être parfaits, montrent toutefois la faisabilité de l'auralisation et que celle-ci repose sur un ensemble de concepts assez simples. Aucun des tests réalisés ne l'a été sur une machine dédiée à cette application, pourtant, les résultats ne nous ont jamais semblés incohérents d'un point de vue perceptif.

Enfin, toutes les personnes ayant écoutés les résultats d'auralisations les ont trouvés plutôt représentatifs. La salle réverbérante a toujours été bien reconnue et le repérage de la source dans l'espace s'est fait sans souci. Aucun test n'a montré une particulièrement bonne estimation de la distance entre source et point d'écoute. A noter enfin que tous ceux ayant écouté les sons résultant, nous compris, ont été surpris du réalisme de certains signaux.

L'auralisation est un champ d'application relativement récent. Même si les premières recherches sur ce thème remontent à plus de 70ans, beaucoup de choses restent à faire. L'utilisation de telles techniques hors des travaux de laboratoire demandera encore beaucoup de travail pour que l'immersion de l'auditoire soit la plus complète possible.

## Annexe A

# Spectres pleine échelle en salle Mersenne

Ce graphique fait écho à la figure V.1

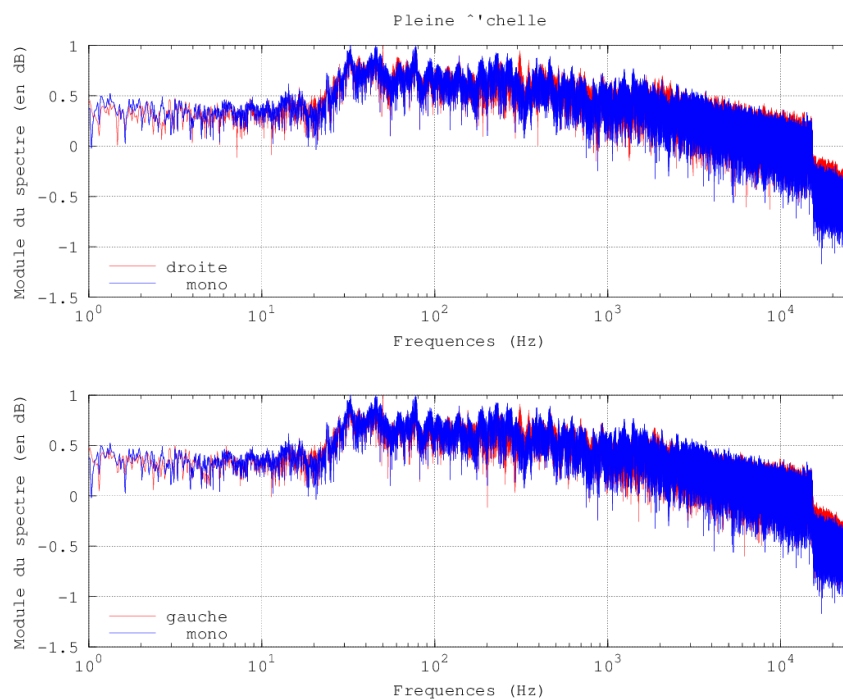


FIGURE A.1 – Les spectres pleine échelle de la comparaison entre les signaux monaural et binaural pour une chanson en salle Mersenne. On note que les enveloppes ne sont pas du tout les mêmes (par rapport au spectre en bleu qui lui, est identique sur les deux figures).

# Bibliographie

- [Bru10] J. Brulez. Auralisation spatialisée de l'effet de salle : synthèse temps réel vs synthèse temps différé. Technical report, Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, 2010.
- [KDS93] M. Kleiner, B. Dalenback, and P. Svensson. Auralization : an overview. *AES Journal*, 41, 1993.
- [Mar12] R. Marchiano. Base du traitement numérique du signal. Technical report, Université Pierre et Marie Curie, Paris VI, 2012.
- [Vor08] M. Vorländer. *Auralization : Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. RWTH edition. Springer London, Limited, 2008.