Université du Maine

Projet auralisation

Comparaison de différents facteurs sur le procédé d'auralisation

L2 SPI

Année universitaire 2012-2013

Thomas Lechat – Xin Wang – Mathieu Gaborit Encadrant : Christophe Ayrault

Sommaire

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 notament la reproduction de transducteurs (en captation ou reproduction).

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 mettre en œuvre une comparaison de l'influence de différents facteurs sur la qualité d'une auralisation, une série de mesures est effectuée (réponses impulsionnelles – RI – binaurales et monaurales, sons en salles cibles, etc...). Ensuite, les signaux mesurés sont convolués avec les RI et le résultat est écouté et qualifié. Finalement, la comparaison même prend forme et les résultats sont consignés et interprétés : il s'agit alors de comparer le résultat obtenu par convolution «*mathématique*» avec le rendu réel (convolution «*physique*» en rejouant le son en salle cible) et ce en variant divers paramètres (RI monaurale/binaurale, mode de convolution, etc...).

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

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

Les deux sont connues depuis longtemps et éprouvés, elles peuvent parallèlement être améliorer de 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 évidentes 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 inexistants en combinant des mesures entre elles ou purement par le calcul. 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 rendu 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 extrèmement complexe et demanderait 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™ [1]

1. CAVE™ : CAVE™ 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 interesser a plusieurs notions qui rentre en jeu dans l'auralisation.

II.1. Produit de convolution

Nous allons nous intéresser dans ce projet uniquement à des salles en 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 :

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

Dans notre cas, si 2 sons d'enveloppes et d'amplitude différentes sont émis dans une salle, il parait logique ques ces sons n'interagiront 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 recépteur 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 $y(t + \tau)$ (voir 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 immuable pendant le temps considéré.

Dans le cadre de ce rapport, à la condition que ni l'émetteur ni le récepteur ne change 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 **??**.

Comme les systèmes qui seront étudiés sont tous linéaires, on peut d'ores et déjà rappeler certaines des propriétés de ces systèmes qui seront utiles par la suite. Sachant

que le système est linéaire, on a donc :

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

Et

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

En combinant ces 2 propriétés on peut en déduire que :

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

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

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

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

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

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

$$e(t) \to s(t) = \int e(\tau) h(t-\tau) d\tau \tag{II. .7}$$

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 ou on prend un i unique). Cette fonction est essentielle dans le processus d'auralisation, il s'agit de la fonction de tranfert caractérisant le système dans le domaine temporelle, 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 interessantes. 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 :

$$\mathcal{F}\{x(t)\} = \int_{-\infty}^{\{} + inftyx(t) e^{-2i\pi Ft} \mathrm{d}x \tag{II. .8}$$

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

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

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

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

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 fait en passant par le domaine de Fourier pour l'accélération (voir le paragraphe III.2.)

II.3. Application à l'auralisation

Dans le cas d'un système linéaire on a :

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

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 que 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 sooient pas parfaites. C'est lors de l'application des compensation pour les sources et chaînes de mesure que passer dans le domaine de Fourier sera réellement utile pour limiter la complexité des calculs de convolution par simple multiplication et soustraction de spectres.

Chapitre III.

Essais et Informations Préliminaires

Avant de chercher à comparer différentes techniques d'auralisation, une étude des moyens de comparaison et divers essais préliminaires sont menés.

III.1. 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. .1) ;

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 du traitement. Afin de conserver les fichiers originaux et d'éviter l'apparition d'incohérences dans les fichiers de mesures, les fenêtrages sont codés en dur 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.

Après deux tests, il apparaît que le fenêtrage des RI apporte un gain non négligeable en terme de temps de calculs [1]. Toutefois, comme le montre la figure III. .2, on remarque que les spectres de signaux convolués avec une RI fenêtrée d'une part et non fenêtrés 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.

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

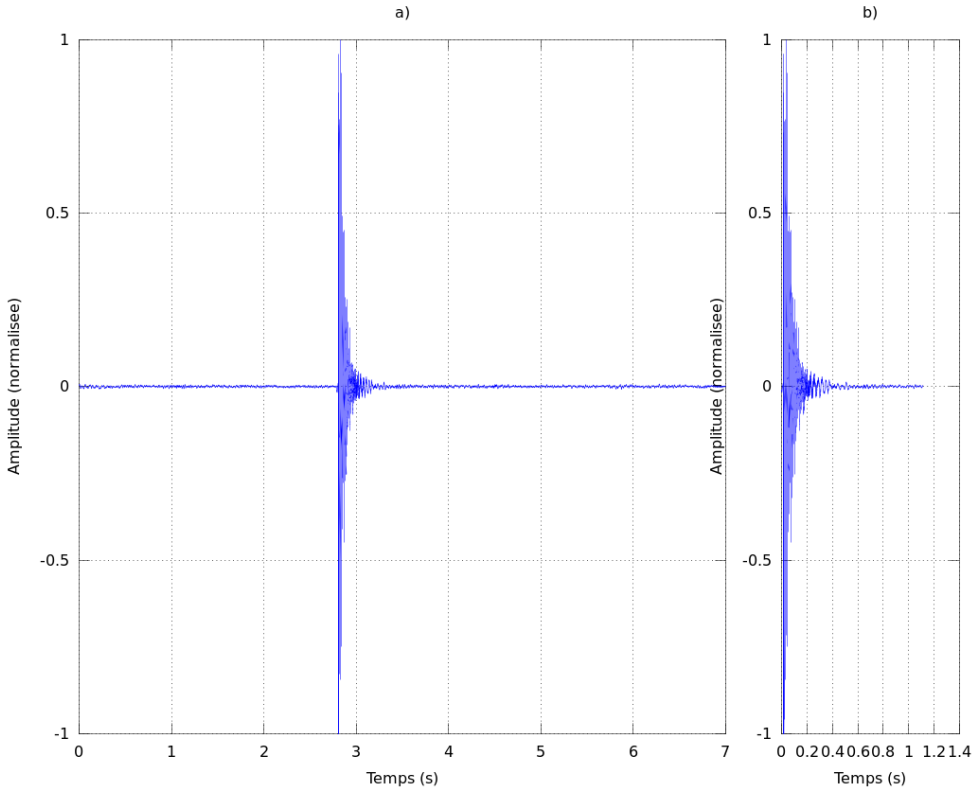

FIGURE III. .1 – Une des RI mesurées avant le fenêtrage (ici, le canal gauche d'une RI binaurale en salle Mersenne). La figure a) montre la RI avant fenêtrage et la figure b) la même RI après fenêtrage

.

III.2. 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 parallèlisée [2], 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. .3).

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

$$TF^{-1}\left\{TF\left\{s(t)\right\}\right\} \approx s(t)$$

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

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

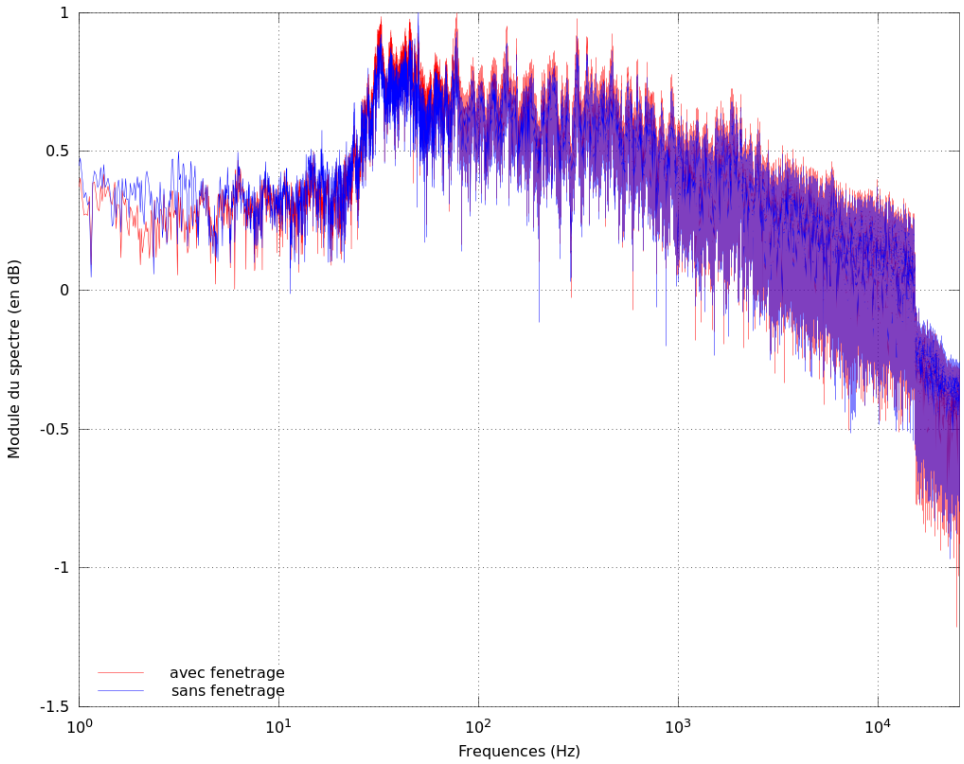

FIGURE III. .2 – Les spectres des sons resultant de la convolution avec une RI fenêtrée (en rouge) et non fenêtrée (en bleu). Même si les écarts sont faibles, ils existent.

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. .3. 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 [3] :

```
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));
```

3. http://www.mathworks.com/matlabcentral/fileexchange/5703-fftconv/content/fftconv.m

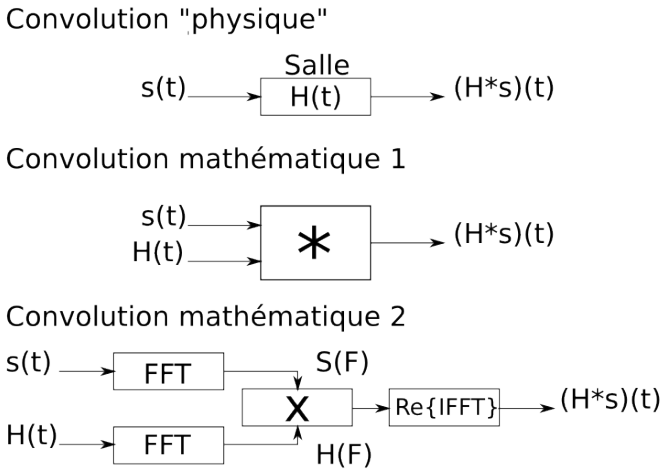

FIGURE III. .3 – Lien entre le projet et la convolution. En haut, la convolution «physique» (lorsque qu'un signal $s(t)$ est émis dans la salle (de réponse impulsionnelle $H(t)$. Au milieu la convolution mathématique au sens strict (*via* l'opération consacrée). En bas, la convolution mathématique en utilisant la propriété de symétrie entre convolution temporelle et multiplication fréquentielle. La chaine du bas est plus longue mais les calculs sont plus rapides que pour celle du milieu.

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

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

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

Le graphe obtenu est présenté en figure III. .4, 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 temps de calcul réalisé, cette fonction semble largement plus avantageuse que `conv()`.

III.3. Utlisation d'un ballon de baudruche

La réponse impulsionnelle d'une salle doit caractériser une salle et particulièrement les différentes réflexion 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.

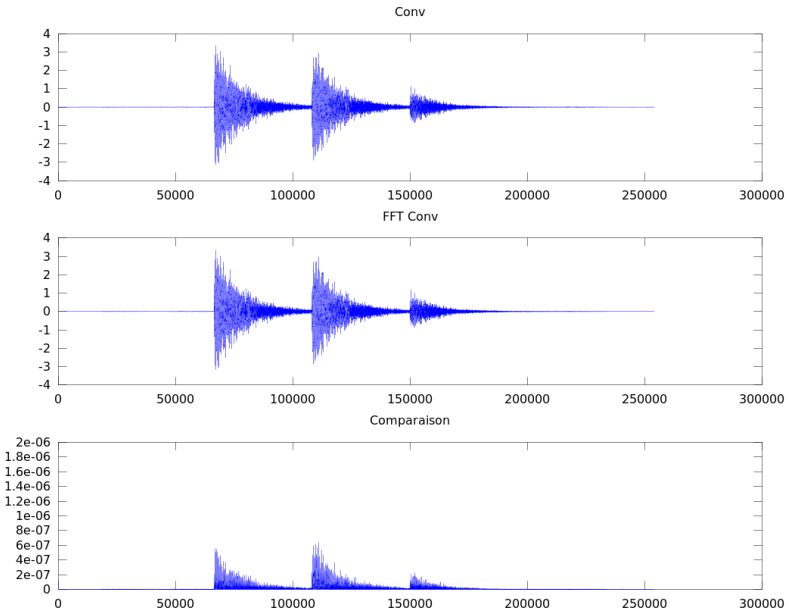

FIGURE III. .4 – Comparaison entre les résultats d'une même convolution avec `conv()` d'une part et `fftconv()` d'autre part. On remarque notamment que sur le graphe du bas (représentant la distance géométrique entre les 2 graphes au dessus), l'axe des ordonnées est gradué entre $2 \cdot 10^{-7}$ et $2 \cdot 10^{-6}$.

L'utilisation d'un ballon de baudruche posait le souci que sa directivité est inconnue. Avec plus de temps, un 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 incovénient soulevé par l'utilisation d'un ballon de baudruche est que la bande passante du son généré par son éclatement n'est pas connu. Un éclatement de ballon de baudruche est donc mesuré en salle semi-anéchoïque (le spectre est visible en figure III. .5.

La réponse en fréquences 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.

Une tentative de compensation des imperfections de la chaine d'excitation est décrite dans la suite.

III.4. 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 si soient suffisament génériques pour ne pas biaiser les résultats mais assez particuliers pour que les altérations produites par la salle cible soient audibles (ou visibles).

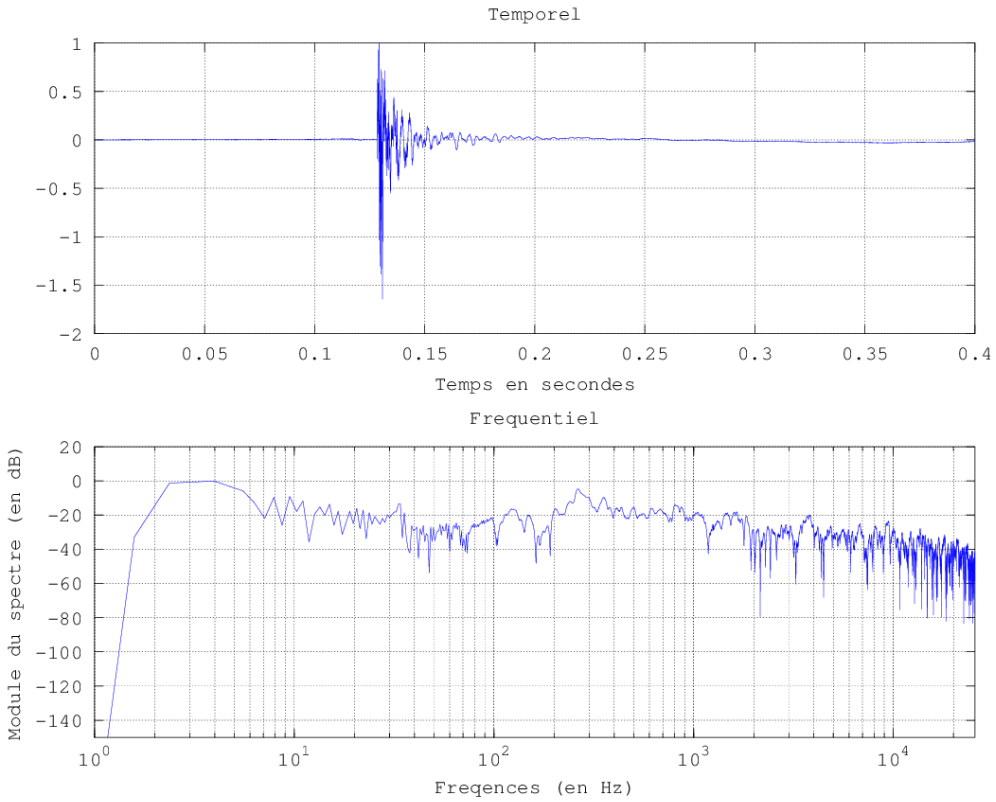

FIGURE III. .5 – Enregistrement temporel de l'éclatement d'un ballon de baudruche en salle semi-anéchoïque (noter la réflexions visible en temporel) et spectre du signal.

D'après Kleiner et coll., la parole ou la musique sont de bons sons de test [KDS93]. le choix des sons de test 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*[4] ; ils ont été enregistrés en salle anéchoïque et échantilonnés à 96 kHz.

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.5. 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 les conditions d'écoute. Deux moyens d'écoute existent (notamment dans le cas d'écoute de résultats binauraux) : au casque ou *via* des HP et un système de restitution.

Si la première méthode ne pose pour ainsi dire aucun problème, la seconde est nette-

4. `http://www.freesound.org/`, merci à *Anton* (`http://www.freesound.org/people/Anton/`) d'avoir posté ces sons là.

ment 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 notament un système anti-diaphonie [KDS93]. De tels systèmes sont parfois difficiles à mettre en place notamment dans 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].

III.6. Salles auralisées

Deux salles ont été utilisées au cours du projet :
- la salle de TP Mersenne (voir figure III. .6) ;
- la salle réverbérante (à coté de la salle mersenne).

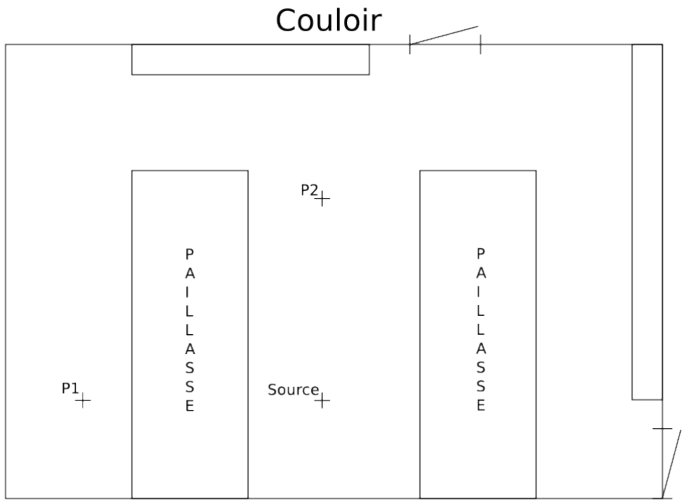

FIGURE III. .6 – Plan de la salle Mersenne. La source a toujours été placée au point noté **Source**. Le micro (pour les prises monaurales) et la tête on toujours étés placés en **P1** ou **P2**. L'orientation par dédfaut de la tête est vers le couloir pour **P1** et vers la source pour **P2**. La source est toujours orientée vers le récepteur.

Chapitre IV.

Comparaison monaural/binaural

La perception sonore humaine est dite binaurale : c'est à dire qu'il y a deux «capteurs» (en l'occurence 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 précsion pouvant aller jusqu'a un degré [Vor08]). Le torse a lui aussi une influence, particulièrement pour le répérage dans le plan vertical. Au cours des mesures pour ce projet, une tête artificielle (sans torse) a été utilisé, le repérage vertical sera donc difficile à reproduire d'après nos mesures.

IV.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. .6) et le recepteur au point **P1** (la tête dans la position par défaut).

IV.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 meilleurs reconnaissance de la géométrie de la salle et des obstacles, en particulier la paillasse présente sur la droite du récepteur.

IV.1.B. Comparaison fréquentielle

On note par ailleurs la différence de contenu fréquentiel (et en particulier la différence de niveau) sur la figure IV. .1. Cette différence observée sur les spectres est bel et bien en accord avec la comparaison perceptive menée ci-avant.

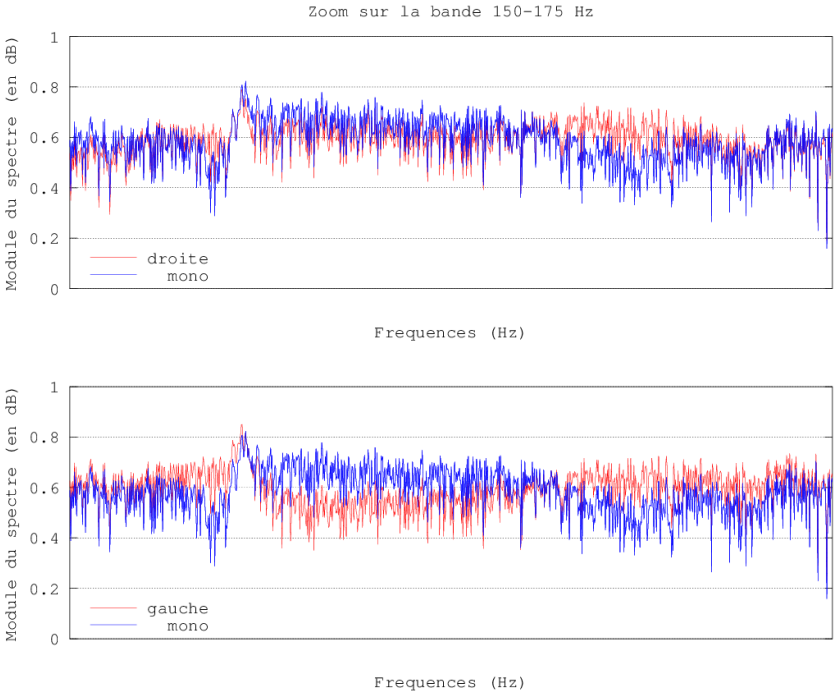

FIGURE IV. .1 – Comparaison entre les contenus fréquentiels des signaux monaural et binaural dans la bande 150-175Hz. On remarque que le signal monaural vient effectivement s'intercaler entre les 2 canaux du signal binaural (en haut monaural et canal droit, en bas monaural et canal gauche). Une représentation pleine échelle est disponible en annexe

Annexe A

Spectres pleine échelle en salle Mersenne

Ce graphique fait écho à la figure IV. .1

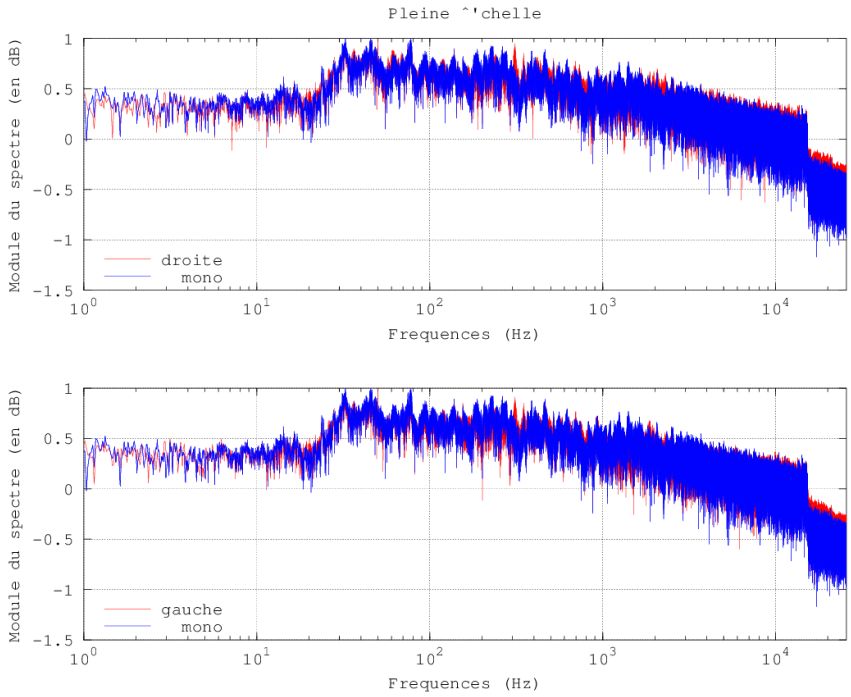

FIGURE A.1 – Les spectres pleine échelle de la comparaison entre les signaux monaural et binaural pour une chanson en salle Mersenne. On note que les enveloppes ne sont pas du tout les mêmes (par rapport au spectre en bleu qui lui, est identique sur les deux figures).

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.

[Vor08] M. Vorländer. *Auralization : Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality.* RWTH edition. Springer London, Limited, 2008.