



HAL
open science

Détection et caractérisation des premières réflexions dans des réponses impulsionnelles spatiales par la fonction d'onde de Herglotz

Pierre Massé, Markus Noisternig

► **To cite this version:**

Pierre Massé, Markus Noisternig. Détection et caractérisation des premières réflexions dans des réponses impulsionnelles spatiales par la fonction d'onde de Herglotz. 16ème Congrès Français d'Acoustique, CFA2022, Société Française d'Acoustique; Laboratoire de Mécanique et d'Acoustique, Apr 2022, Marseille, France. hal-03848221

HAL Id: hal-03848221

<https://hal.science/hal-03848221>

Submitted on 10 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



16^{ème} Congrès Français d'Acoustique
11-15 Avril 2022, Marseille

Détection et caractérisation des premières réflexions dans des réponses impulsionnelles spatiales par la fonction d'onde de Herglotz

P. Massé ^a, M. Noisternig ^a

^a Équipe Espaces Acoustiques et Cognitifs, STMS, Ircam – Sorbonne Université – CNRS – Ministère de la Culture



La mesure de réponses impulsionnelles spatiales de salles par des antennes sphériques de microphones permet la reproduction tri-dimensionnelle de l'effet de réverbération de salle par convolution multicanal sur des dispositifs sonores immersifs (“*surround-sound*”). L'avantage de cette approche est que la réponse impulsionnelle représente l'empreinte acoustique exacte de l'espace mesuré ; cependant, une mesure donnée demeure statique par nature. De plus, l'opération de convolution multicanal est très coûteuse en termes de puissance de calcul numérique. Certaines méthodes paramétriques cherchent à contourner ces problèmes en décomposant la réponse impulsionnelle trame par trame en une composante entièrement directionnelle et une autre parfaitement diffuse. Les réverbérateurs hybrides cherchent plutôt à alléger le coût de calcul en reproduisant la queue de réverbération par un réseau de retards bouclés, limitant la convolution aux réflexions précoces. Une analyse de la réponse mesurée permet de modéliser la queue de réverbération tardive ainsi paramétrer correctement le réseau de retards. Les caractéristiques de la réverbération tardive peuvent ensuite être facilement modifiées en manipulant les paramètres du réseau. Une modélisation des premières réflexions permettrait quant à elle de modifier la partie précoce de la réponse, ouvrant ainsi la voie à une manipulation complète de l'effet de réverbération.

Dans cette communication, nous proposons une méthode de détection et de caractérisation des premières réflexions basée sur le formalisme de la fonction d'onde de Herglotz. Cette approche est évaluée sur des signaux simulés et comparée à une localisation plus classique par formation de voies. Enfin, des exemples d'application à des mesures réelles sont présentés et une brève discussion est ouverte sur les différentes stratégies de manipulation rendues possibles par ce travail.

1 Introduction

Cette communication s'intéresse à la détection et à la caractérisation des premières réflexions dans des réponses impulsionnelles spatiales de salles (SRIR). Nous présentons ici une approche basée sur le formalisme de la fonction d'onde de Herglotz, qui permet de décrire le champ mesuré par une antenne sphérique de microphones (SMA) comme une superposition d'ondes planes munies d'un noyau (*Herglotz kernel*) représentant la probabilité de la présence de l'onde sur la surface du SMA.

Ce travail s'inscrit dans le contexte du développement d'un cadre d'analyse-traitement-manipulation complet de SRIR. L'objectif de cette boîte à outils est de permettre l'analyse d'une SRIR mesurée par un SMA à des fins soit de reproduction par des techniques de réverbération artificielle, soit de modification des caractéristiques temps-fréquence-espace de la SRIR, ou soit encore de piloter des approches “hybrides” mêlant les deux.

Dans ce cadre, la caractérisation des premières réflexions passe alors par une localisation trame-par-trame des échos discrets prédominants contenus dans la partie communément appelée “précoce” d'une SRIR.

2 Bases théoriques

2.1 Segmentation temporelle de SRIR

Nous considérons ici qu'une SRIR peut être segmentée temporellement en une partie “précoce” contenant principalement les premières réflexions discrètes, à la fois en temps et en espace, et une autre “tardive” contenant exclusivement la queue de réverbération stochastique et exponentiellement décroissante. Cette segmentation s'opère à un temps de transition souvent dénoté t_{mix} , relatif au temps d'arrivée du son direct (t_0). Le présent travail s'intéresse donc aux trames temporelles contenues entre t_0 et t_{mix} .

2.2 Champ mesuré par un SMA

Une onde plane incidente d'une direction $\mathbf{\Omega}_d$ donnée et d'amplitude unitaire peut être entièrement décrite dans le domaine des harmoniques sphériques (SH) selon [1, p. 227] :

$$X_{\text{in}}(k, \mathbf{r}, \mathbf{d}) = e^{i\mathbf{k}\cdot\mathbf{r}\cdot\mathbf{d}}$$

$$X_{\text{in}}(kr, \mathbf{\Omega}, \mathbf{\Omega}_d) = 4\pi \sum_{l=0}^{\infty} i^l j_l(kr) \sum_{m=-l}^l Y_{l,m}(\mathbf{\Omega}) Y_{l,m}^*(\mathbf{\Omega}_d), \quad (1)$$

où $\mathbf{\Omega} = [\theta, \varphi]$ est la position angulaire en azimuth-élévation, $i = \sqrt{-1}$ est le nombre imaginaire, $k = \omega/c$ est le nombre d'onde de l'onde plane (avec sa fréquence angulaire $\omega = 2\pi f$ et la célérité de l'air c), et $j_l(kr)$ sont les fonctions de Bessel sphériques du premier type.

Dans le cas d'une sphère rigide de rayon $r = r_s$, ce qui représente la majorité des SMA disponibles au commerce aujourd'hui (p. ex. le Eigenmike[®] de mh acoustics), le champ dispersé doit aussi être pris en compte [1, p. 228] :

$$X_{\text{sc}}(kr, \mathbf{\Omega}, \mathbf{\Omega}_d) = -4\pi \sum_{l=0}^{\infty} i^l \frac{h'_l(kr_s)}{j'_l(kr_s)} h_l(kr) \sum_{m=-l}^l Y_{l,m}(\mathbf{\Omega}) Y_{l,m}^*(\mathbf{\Omega}_d), \quad (2)$$

où $h_l(kr)$ sont les fonctions de Hankel sphériques du premier type et $'$ représente ici une dérivation par rapport à l'argument kr . Le champ sonore complet généré par une onde plane incidente sur la surface d'un SMA rigide est alors donné par :

$$X_{\text{PW}}(kr, \mathbf{\Omega}, \mathbf{\Omega}_d) = X_{\text{in}}(kr, \mathbf{\Omega}, \mathbf{\Omega}_d) + X_{\text{sc}}(kr, \mathbf{\Omega}, \mathbf{\Omega}_d)$$

$$= \sum_{l=0}^{\infty} b_l(kr) \sum_{m=-l}^l Y_{l,m}(\mathbf{\Omega}) Y_{l,m}^*(\mathbf{\Omega}_d), \quad (3)$$

où $b_l(kr) = 4\pi i^l \left[j_l(kr) - \frac{h'_l(kr_s)}{j'_l(kr_s)} h_l(kr) \right]$ est souvent appelée la *fonction holographique* d'une sphère rigide.

Nous nous intéressons ici plus particulièrement au champ mesuré par les transducteurs du SMA, c'est à dire à des positions discrètes $\mathbf{r}_q = [r_s, \mathbf{\Omega}_q]$ réparties sur

sa surface. La théorie de la conception de SMA et de la “bonne” répartition de ces points de mesure sur la sphère est au-delà des considérations du présent article ; les lecteurs les plus curieux sont invités à consulter la littérature “classique” en la matière, et notamment le travail fondateur de Rafaely [2, 3]. Supposant alors que le SMA est pour ainsi dire “bien conçu”, la mesure du champ sonore aux positions $\mathbf{r}_q = (r_s, \mathbf{\Omega}_q)$ permet alors de prendre sa transformée en SH (discrète et tronquée) :

$$X_{l,m}(r_s) = \sum_q \alpha_q X(r_s, \mathbf{\Omega}_q) Y_{l,m}^*(\mathbf{\Omega}_q), \quad (4)$$

où α_q sont des poids de quadrature optimisés afin de minimiser les erreurs de discrétisation et de troncature. Notons au passage que l'ordre maximal L atteignable par cette transformée en SH est limité par le nombre de valeurs singulières non-nulles présentées par la matrice d'encodage $E_{n,q} = \alpha_q Y_{l,m}^*(\mathbf{\Omega}_q)$, où $n = l^2 + l + m$ [8, 9]. Dans le cas de l'Eigenmike et de ses $Q = 32$ transducteurs, cela donne un ordre maximal $L = 4$.

Dans le cas d'une décomposition en ondes planes (*plane wave decomposition*, PWD), c'est-à-dire en supposant que le champ mesuré est une superposition d'ondes planes telles que décrites par l'Eq. (3), la transformée en SH permet de simplifier l'expression du champ sonore grâce à l'orthogonalité des harmoniques sphériques :

$$\begin{aligned} X_{l,m}(kr_s, \mathbf{\Omega}_d) &= \sum_q \alpha_q X_{\text{PW}}(kr_s, \mathbf{\Omega}_q, \mathbf{\Omega}_d) Y_{l,m}^*(\mathbf{\Omega}_q) \\ &= \sum_{l'=0}^L b_{l'}(kr_s) \sum_{m'=-l'}^{l'} Y_{l',m'}^*(\mathbf{\Omega}_d) \\ &\quad \times \sum_q \alpha_q Y_{l',m'}(\mathbf{\Omega}_q) Y_{l,m}^*(\mathbf{\Omega}_q) \\ &\approx \sum_{l'=0}^L b_{l'}(kr_s) \sum_{m'=-l'}^{l'} Y_{l',m'}^*(\mathbf{\Omega}_d) \delta_{l,l'} \delta_{m,m'} \\ &= b_l(kr_s) Y_{l,m}^*(\mathbf{\Omega}_d). \end{aligned} \quad (5)$$

Remarquons ici que le produit kr_s , étant fixé à la surface du SMA, n'exprime plus qu'une simple dépendance en fréquence $f = kc/2\pi$. De plus, si la fonction holographique $b_l(f)$ de la sphère peut ensuite être compensée, le champ résultant [disons $\tilde{X}_{l,m}(f, \mathbf{\Omega})$] serait alors indépendant du SMA par lequel il fût mesuré.

En effet, cette idée est une des principales motivations derrière le formalisme *Higher-Order Ambisonics* (HOA), tel qu'initialement conçu par Daniel [4]. La compensation de la fonction holographique se fait alors par l'application de filtres “correcteurs” [5] dont les détails ne sont pas pertinents ici mais dont la réponse en fréquence $a_l(f)$ donne au champ résultant $\tilde{X}_{l,m}(f, \mathbf{\Omega})$ une réponse en fréquence $c_l(f) = a_l(f)b_l(f)$. Les stratégies de conception des filtres correcteurs cherchent donc, pour la plupart, à générer une réponse $c_l(f)$ la plus “plate” et la plus indépendante des caractéristiques de $b_l(f)$ possible.

Cependant, pour différentes raisons détaillées (entre autres) par Daniel et Moreau [5], la réponse en fréquence

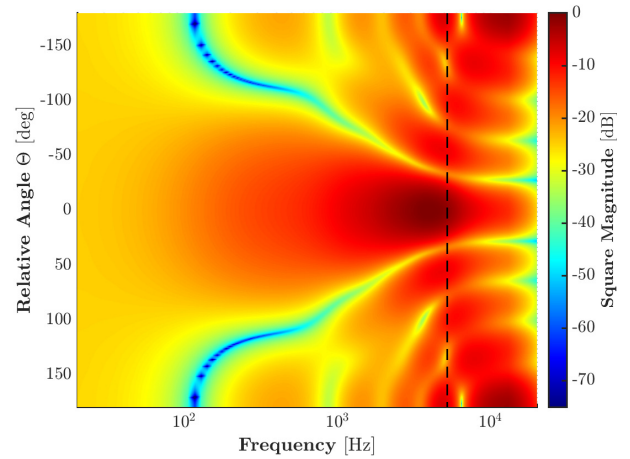


FIGURE 1 – Fonction de directivité $w(f, \mathbf{\Omega})$ sur le plan azimuthal pour l'encodage en *Higher-Order Ambisonics* (HOA) à l'ordre 4 d'une onde plane mesurée par l'Eigenmike de mh acoustics. La ligne hachée verticale (noire) indique la fréquence de repliement spatial f_{alias} .

résultante $c_l(f)$ n'est jamais parfaitement plate. Par conséquent, la *fonction de directivité* de cette PWD “encodée” en HOA présente elle aussi une dépendance fréquentielle :

$$w(f, \mathbf{\Omega}) = \sum_{l=0}^L \sum_{m=-l}^l c_l(f) Y_{l,m}(\mathbf{\Omega}) Y_{l,m}^*(\mathbf{\Omega}_d). \quad (6)$$

La Figure 1 illustre cette fonction de directivité $w(f, \mathbf{\Omega})$ sur le plan azimuthal pour l'encodage en HOA à l'ordre $L = 4$ d'une onde plane mesurée par l'Eigenmike de mh acoustics, mentionné ci-dessus.

En calculant l'indice de directivité (*directivity index*, DI) [6], nous pouvons constater qu'il existe une bande de directivité maximale ayant pour borne supérieure la fréquence de repliement spatial f_{alias} [7]. Cette bande de directivité maximale est illustrée à la Figure 2 en prenant l'Eigenmike une nouvelle fois à titre d'exemple.

2.3 Fonction d'onde de Herglotz

Nous supposons (voir la Section 2.1 ci-dessus) que les réflexions précoces prédominantes se comportent (plus ou moins) comme des impulsions. Dans le contexte d'un champ sonore mesuré par un SMA, cela peut être vu comme une superposition d'ondes planes couvrant toutes les fréquences et incidentes simultanément. Ainsi, le formalisme de la fonction d'onde de Herglotz [10, p. 60] paraît particulièrement adapté. Essentiellement, la fonction d'onde de Herglotz décrit la présence d'une onde plane incidente sur une surface donnée à travers une fonction de probabilité appelée le noyau (*kernel*) de Herglotz.

Le noyau de Herglotz est une fonction continue (au moins C^1) et de carré intégrable (c.à.d. L^2) sur la surface donnée, et donc le problème de la localisation de la position de l'onde plane incidente sur la surface peut être considéré comme un

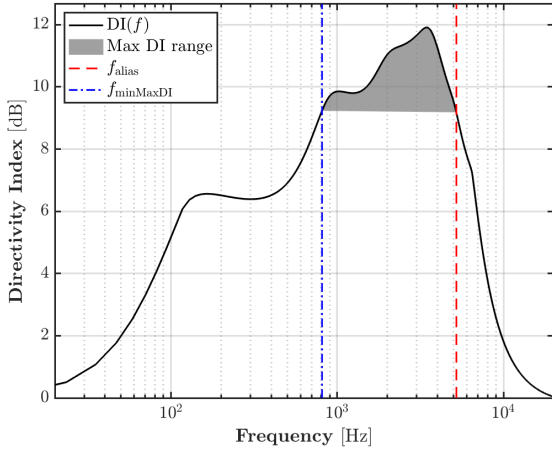


FIGURE 2 – Indice de directivité (DI) en fonction de la fréquence pour l'encodage en *Higher-Order Ambisonics* (HOA) à l'ordre 4 d'une onde plane mesurée par l'Eigenmike de mh acoustics. La ligne hachée verticale (rouge) indique la fréquence de repliement spatial f_{alias} , et la zone grisée représente une bande de directivité maximale telle que $\text{DI}(f_b) \geq \text{DI}(f_{\text{alias}}) \forall f_b \in [f_{\text{minMaxDI}}, f_{\text{alias}}]$.

problème de détection de pic sur le noyau. Dans le cas d'une surface sphérique S_r^2 , la fonction d'onde de Herglotz $v(k, \mathbf{r})$ peut être écrite ainsi :

$$v(k, \mathbf{r}) = \int_{S_r^2} g(\boldsymbol{\Omega}_d) e^{i\mathbf{k}\mathbf{r}\cdot\mathbf{d}} d\boldsymbol{\Omega}_d, \quad (7)$$

où $\mathbf{d} = [1, \boldsymbol{\Omega}_d]$ et $g(\boldsymbol{\Omega}_d)$ est le noyau de Herglotz.

La méthode que nous présentons ici consiste alors à décrire la mesure de réflexions précoces par un SMA comme une superposition de fonctions d'ondes de Herglotz et ainsi détecter leurs directions d'arrivée (*directions of arrival*, DoA) à partir d'une estimation de leurs noyaux.

3 Méthode proposée

3.1 Fonction de localisation

Colton et Kress [10, p. 68] démontrent que si l'on décrit une onde plane incidente sur une sphère rigide sous la forme d'une fonction d'onde de Herglotz :

$$\begin{aligned} v_{\text{in}}(k, \mathbf{r}_s) &= \int_{S_r^2} g(\boldsymbol{\Omega}_d) e^{i\mathbf{k}\mathbf{r}_s\cdot\mathbf{d}} d\boldsymbol{\Omega}_d \\ v_{\text{in}}(kr_s, \boldsymbol{\Omega}) &= \int_{S_r^2} g(\boldsymbol{\Omega}_d) X_{\text{in}}(kr_s, \boldsymbol{\Omega}, \boldsymbol{\Omega}_d) d\boldsymbol{\Omega}_d, \end{aligned} \quad (8)$$

alors le champ complet (incluant les effets de dispersion de la sphère) peut être décrit de la même manière :

$$\begin{aligned} v(kr_s, \boldsymbol{\Omega}) &= \int_{S_r^2} g(\boldsymbol{\Omega}_d) X_{\text{PW}}(kr_s, \boldsymbol{\Omega}, \boldsymbol{\Omega}_d) d\boldsymbol{\Omega}_d \\ &= \sum_{l=0}^{\infty} b_l(kr_s) \sum_{m=-l}^l Y_{l,m}(\boldsymbol{\Omega}) \int_{S_r^2} g(\boldsymbol{\Omega}_d) Y_{l,m}^*(\boldsymbol{\Omega}_d) d\boldsymbol{\Omega}_d. \end{aligned} \quad (9)$$

À l'instar de l'Eq. (5), cette expression peut aussi être simplifiée en prenant sa transformée en SH :

$$\begin{aligned} v_{l,m}(f) &= \sum_{l'=0}^L b_{l'}(f) \sum_{m'=-l'}^{l'} \sum_q \alpha_q Y_{l',m'}(\boldsymbol{\Omega}_q) Y_{l,m}^*(\boldsymbol{\Omega}_q) \\ &\quad \times \int_{S_r^2} g(\boldsymbol{\Omega}_d) Y_{l',m'}(\boldsymbol{\Omega}_d) d\boldsymbol{\Omega}_d \\ &\approx \sum_{l'=0}^L b_{l'}(f) \sum_{m'=-l'}^{l'} \delta_{l,l'} \delta_{m,m'} \int_{S_r^2} g(\boldsymbol{\Omega}_d) Y_{l',m'}^*(\boldsymbol{\Omega}_d) d\boldsymbol{\Omega}_d \\ &= b_l(f) \int_{S_r^2} g(\boldsymbol{\Omega}_d) Y_{l,m}^*(\boldsymbol{\Omega}_d) d\boldsymbol{\Omega}_d. \end{aligned} \quad (10)$$

Puisque nous cherchons à estimer numériquement le noyau $g(\boldsymbol{\Omega}_d)$, il est alors nécessaire de discrétiser l'intégrale de la fonction d'onde de Herglotz. De plus, puisque nous cherchons à travailler sur des SRIR encodées en HOA, nous considérons les signaux filtrés présentant la réponse en fréquence $c_l(f)$ à la place de la fonction holographique $b_l(f)$. Cela nous permet enfin d'écrire le problème sous forme matricielle :

$$\mathbf{v}_{\text{HOA}}(f) = \mathbf{D}_{\text{HOA}}(f) \mathbf{g}, \quad (11)$$

où $D_{n,d} = \beta_d c_l(f) Y_{l,m}^*(\boldsymbol{\Omega}_d)$ est une matrice de taille $(L+1)^2 \times D$ (β_d sont les poids de quadrature pour la grille sphérique $\boldsymbol{\Omega}_d$) et $\mathbf{g} = [g(\boldsymbol{\Omega}_0), g(\boldsymbol{\Omega}_1), \dots, g(\boldsymbol{\Omega}_d), \dots, g(\boldsymbol{\Omega}_{D-1})]^T$.

Considérons maintenant la superposition de P fonctions d'ondes de Herglotz simultanées (c.à.d. présentes au sein de la même trame d'analyse), chacune possédant sa propre amplitude relative $a_p(f)$ et son propre noyau $g_p(\boldsymbol{\Omega}_d)$:

$$\tilde{v}_{l,m}(f) = c_l(f) \sum_d \alpha_d Y_{l,m}^*(\boldsymbol{\Omega}_d) \sum_p a_p(f) g_p(\boldsymbol{\Omega}_d), \quad (12)$$

ou sous forme matricielle :

$$\begin{aligned} \mathbf{v}_{\text{HOA}}(f) &= \mathbf{D}_{\text{HOA}}(f) \mathbf{G} \mathbf{a}(f) \\ &= \mathbf{D}_{\text{HOA}}(f) \mathbf{m}(f), \end{aligned} \quad (13)$$

où $\mathbf{G} = [\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_p, \dots, \mathbf{g}_{P-1}]$, $\mathbf{a}(f)$ contient les amplitudes $a_p(f)$, et $\mathbf{m}(f) = \mathbf{G} \mathbf{a}(f)$ représente ce que l'on appellera la *fonction de localisation*.

Il s'agit alors d'estimer cette fonction de localisation $\mathbf{m}(f)$ à partir d'une trame $\mathbf{v}_{\text{HOA}}(f)$ mesurée. Suivant l'Eq. (13), cela peut notamment se faire par une inversion de la matrice $\mathbf{D}_{\text{HOA}}(f)$:

$$\tilde{\mathbf{m}}(f) = \mathbf{D}_{\text{HOA}}(f)^{-1} \mathbf{v}_{\text{HOA}}(f). \quad (14)$$

Cependant, une telle inversion (ou pseudo-inversion) serait inévitablement limitée dans sa dimensionalité : puisque nous savons que notre représentation HOA est de taille $(L + 1)^2$, le “meilleur” cas possible pour une inversion directe serait celui d’une matrice $\mathbf{D}_{\text{HOA}}(f)$ carrée, c’est-à-dire $D = (L + 1)^2$. Afin d’obtenir une meilleure résolution angulaire, nous sommes alors confronté à un problème sous-déterminé, avec $D > (L + 1)^2$.

De ce fait, nous proposons d’exploiter la méthode de régularisation dite de Tikhonov et d’estimer la fonction de localisation de la manière suivante :

$$\tilde{\mathbf{m}}(f) = \left[\mathbf{D}_{\text{HOA}}^\top(f) \mathbf{D}_{\text{HOA}}(f) + \mathbf{T}^\top \mathbf{T} \right]^{-1} \mathbf{D}_{\text{HOA}}^\top(f) \mathbf{v}_{\text{HOA}}(f), \quad (15)$$

où $\mathbf{T} = \mu \mathbf{I}$ est une matrice de Tikhonov (ou plus techniquement un opérateur de régularisation) favorisant les solutions à petites normes, avec μ le paramètre de régularisation et \mathbf{I} la matrice identité de taille $D \times D$.

Afin de simplifier les notations, nous avons considéré jusqu’ici les fréquences f de manière continue ; or, dans une analyse trame-par-trame (p. ex. par transformée de Fourier à court terme discrète), celles-ci sont invariablement discrétisées en K fenêtres (*bins*) f_k recouvrant les fréquences positives. Nous pouvons donc, en théorie, estimer la fonction de localisation selon l’Eq. (15) à chaque fenêtre f_k . Cependant, nous avons vu à la Section 2.2 que la directivité d’une onde plane encodée en HOA évolue fortement avec la fréquence. Au-delà de la fréquence de repliement spatial f_{alias} , notamment, l’information directionnelle devient ambiguë, et il est donc inutile d’essayer de définir une fonction de localisation à ces fréquences. Il convient donc de restreindre le calcul de la fonction de localisation à la bande de directivité maximale illustrée à la Figure 2, c’est-à-dire $f_k \in [f_{\text{minMaxDI}}, f_{\text{alias}}]$ tel que $\text{DI}(f_k) \geq \text{DI}(f_{\text{alias}})$.

3.2 Détection d’échos

Nous cherchons maintenant à établir une stratégie de détection de pics sur la fonction de localisation $\tilde{\mathbf{m}}(f_k)$, ce qui correspondra *in fine* à une estimation des directions d’arrivée (DoA) des échos incidents. À noter que, de par la construction de $\tilde{\mathbf{m}}(f_k)$, cela implique aussi une caractérisation de leurs amplitudes relatives.

Remarquons d’abord que la fonction de localisation $\tilde{\mathbf{m}}(f_k)$ est, de manière générale, complexe ; or, l’information directement liée à la “présence” (ou non) d’une réflexion est seulement portée par son module. De plus, rappelons que les noyaux de Herglotz $g_p(\boldsymbol{\Omega}_d)$ sont indépendants de la fréquence : notre stratégie de détection de pics devrait alors l’être aussi. Nous proposons donc de travailler sur une moyenne pondérée du module de la fonction de localisation sur les fréquences d’intérêt $f_k \in [f_{\text{minMaxDI}}, f_{\text{alias}}]$:

$$\bar{m}_d = \frac{1}{\Gamma_{\text{tot}}} \sum_k \gamma_k |\tilde{m}_{d,k}|, \quad (16)$$

où $\gamma_k = \max_d (\tilde{m}_{d,k})$ et $\Gamma_{\text{tot}} = \sum_k \gamma_k$.

La fonction de localisation “finale” $\bar{\mathbf{m}}$ ainsi définie, il devient alors possible de développer une stratégie de détection des P pics directionnels présentés par celle-ci. Ici,

nous faisons le choix de modéliser directement la fonction de localisation par une superposition de P fonctions de bases radiales (*radial basis functions*, RBF). Plus particulièrement, nous utilisons des gaussiennes radiales :

$$\widehat{m}_d(\mathbf{d}_p, h_p, \lambda_p) = \sum_p \sqrt{h_p^2} e^{\lambda_p(\mathbf{d}_p \cdot \mathbf{r}_d - 1)}, \quad (17)$$

où $\mathbf{d}_p = [1, \boldsymbol{\Omega}_p]$ est la position du pic de chaque gaussienne sur la sphère, h_p est la “présence” de chacune, λ_p et le paramètre de “raideur” du pic de la gaussienne, et $\mathbf{r}_d = [1, \boldsymbol{\Omega}_d]$. La détection de pics devient alors une estimation des paramètres \mathbf{d}_p , qui donne alors le DoA $\boldsymbol{\Omega}_d$ de chaque réflexion incidente, et h_p , leur présences respectives (ce qui donne une certaine caractérisation de leurs amplitudes relatives).

Nous pouvons à nouveau écrire le problème sous forme matricielle :

$$\widehat{\mathbf{m}}(\boldsymbol{\Psi}) = \mathbf{M} \mathbf{h}, \quad (18)$$

où \mathbf{h} contient les présences $\sqrt{h_p^2}$, $M_{d,p} = e^{\lambda_p(\mathbf{d}_p \cdot \mathbf{r}_d - 1)}$, et la matrice $\boldsymbol{\Psi}$ regroupe les paramètres à estimer $\boldsymbol{\psi} = [\mathbf{d}_p^\top, h_p, \lambda_p]^\top$. À noter que le paramètre de présence est délibérément construit sous la forme $\sqrt{h_p^2}$ afin d’assurer que celui-ci reste non-négatif au cours de la modélisation.

Enfin, nous choisissons d’estimer les paramètres \mathbf{d}_p , h_p , et λ_p en opérant une descente de gradient sur le problème linéaire décrit par l’Eq. (18). Cette descente de gradient cherche à minimiser la fonction objective (ou “de coût”) suivante :

$$\mathcal{L}(\boldsymbol{\Psi}) = \frac{1}{M_{\text{tot}}} \sum_d \bar{m}_d [\bar{m}_d - \widehat{m}_d(\boldsymbol{\Psi})]^2, \quad (19)$$

où $M_{\text{tot}} = \sum_d \bar{m}_d$. Cette fonction est essentiellement une erreur moyenne carrée (*mean squared error*, MSE) classique à laquelle nous avons rajouté une pondération proportionnelle à la fonction de localisation $\bar{\mathbf{m}}$. En d’autres mots, l’erreur dans la modélisation $\mathbf{m}(\boldsymbol{\Psi})$ est évaluée de manière prédominante dans les directions où la fonction de localisation est importante, ce qui découle du fait que l’on cherche principalement à assurer une bonne localisation des pics. Cela permet aussi, par ailleurs, d’éviter de prendre en compte un éventuel bruit de fond présent dans les régions de la sphère ne contenant aucun pic.

3.3 Estimation du nombre d’échos incidents

Nous avons supposé jusqu’ici que le nombre P de pics à détecter était une valeur “connue”. Or il est évident que cela n’est pas le cas *a priori* au moment d’analyser une trame de SRIR donnée. Nous utilisons ici une estimation inspirée des approches ESPRIT dans le domaine SH [11] et basée sur une décomposition en éléments propres de la matrice de covariance (normalisée) :

$$R_{n,n'} = \frac{1}{K_b} \sum_k \frac{\tilde{v}_{l,m}(f_k) \tilde{v}_{l',m'}^*(f_k)}{\tilde{v}_{l,m} \tilde{v}_{l',m'}}, \quad (20)$$

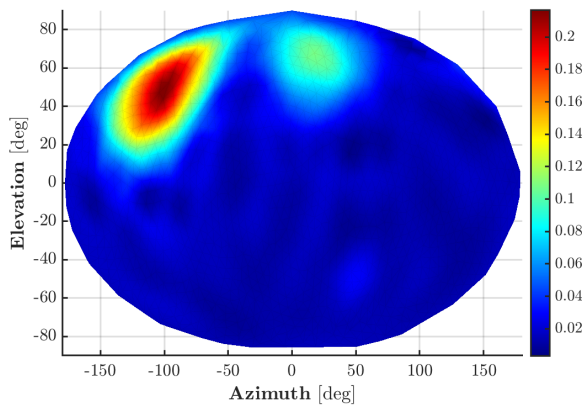


FIGURE 3 – Fonction de localisation $\tilde{\mathbf{m}}$, Eq. (16), pour une trame contenant $P = 2$ échos placés à $\mathbf{\Omega}_1 = [-120^\circ, 45^\circ]$ et $\mathbf{\Omega}_2 = [22.5^\circ, 60^\circ]$, respectivement. La deuxième impulsion est synthétisée à -6 dB par rapport à la première.

où $\tilde{v}_{l,m} = \sum_k |\tilde{v}_{l,m}(f_k)|^2 / K_b$ et K_b est le nombre de fenêtres fréquentielles dans la bande d'intérêt.

Le nombre d'échos présents dans la trame donnée est ensuite proportionnel au nombre de valeurs propres "non-nulles". En pratique, puisque nous sommes limités à un calcul de la matrice de covariance sur la seule trame d'analyse, il convient de définir un seuil $s_{\text{thresh}} = \bar{s} + s_\sigma$, où s sont les valeurs propres de \mathbf{R} , \bar{s} leur moyenne, et s_σ leur écart-type. Le nombre de valeurs propres vérifiant $s > s_{\text{thresh}}$ donne alors notre estimation de P .

4 Signaux d'évaluation

Nous présentons ici une évaluation très fondamentale sur des signaux synthétiques construits afin de simuler une trame d'une SRIR mesurée avec l'Eigenmike. Ces signaux de synthèse contiennent $P = 2$ impulsions filtrées entre 125 Hz et 16 kHz et placées à $\mathbf{\Omega}_1 = [-120^\circ, 45^\circ]$ et $\mathbf{\Omega}_2 = [22.5^\circ, 60^\circ]$, respectivement, et de manière à se trouver toutes les deux sur une seule trame centrale de 256 échantillons. La deuxième impulsion est affectée d'une amplitude relative de -6 dB par rapport à la première. Un bruit de fond gaussien est ajouté à une puissance de -80 dB.

5 Résultats

5.1 Fonction de localisation

La Figure 3 montre la fonction de localisation $\tilde{\mathbf{m}}$ telle qu'obtenue à partir des Eqs. (15) et (16). À noter que la fonction de localisation $\tilde{\mathbf{m}}(f_k)$ de l'Eq. (15) est calculée sur une grille $\mathbf{\Omega}_d$ de type Sloan-Womersley [12] de taille $D_1 = 81$ et que $\tilde{\mathbf{m}}$ est ensuite interpolée sur une deuxième grille Sloan-Womersley de taille $D_2 = 361$ par interpolation

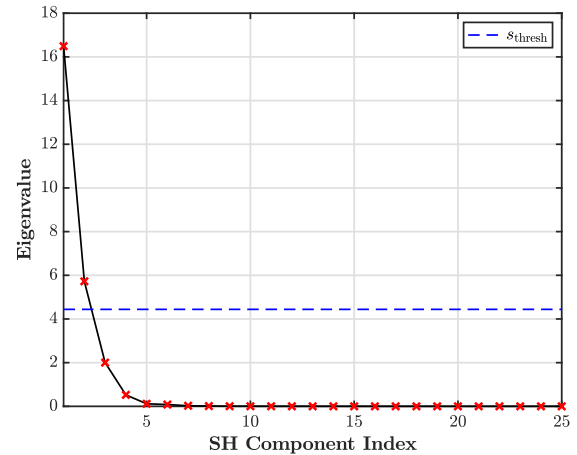


FIGURE 4 – Spectre des valeurs propres de la matrice de covariance normalisée \mathbf{R} , Eq. (20), pour une trame contenant $P = 2$ échos placés à $\mathbf{\Omega}_1 = [-120^\circ, 45^\circ]$ et $\mathbf{\Omega}_2 = [22.5^\circ, 60^\circ]$, respectivement. La ligne horizontale hachée (bleue) indique le seuil s_{thresh} utilisé pour estimer le nombre de réflexions présentes dans la trame (P).

Hermite cubique sphérique [13]. Enfin, pour obtenir une visualisation plus lisse, la Figure 3 est à nouveau interpolée sur une grille de taille $D_3 = 1369$.

Le paramètre de régularisation de Tikhonov μ est estimé par validation croisée généralisée (*generalized cross-validation*, GCV) avec une contrainte additionnelle sur le conditionnement de la matrice $\mathbf{D}_{\text{HOA}}^\top(f_k)\mathbf{D}_{\text{HOA}}(f_k) + \mathbf{T}^\top\mathbf{T}$ résultante afin de favoriser des solutions "lisses".

5.2 Estimation du nombre d'échos

La Figure 4 montre le spectre des valeurs propres de la matrice de covariance normalisée \mathbf{R} telle qu'obtenue selon l'Eq. (20), ainsi que le seuil $s_{\text{thresh}} = \bar{s} + s_\sigma$ utilisé pour estimer le nombre de réflexions P présentes dans la trame (\bar{s} étant la moyenne des valeurs propres et s_σ leur écart-type).

5.3 Détection de pics

Enfin, la Figure 5 montre la fonction de localisation reconstruite $\tilde{\mathbf{m}}(\Psi)$ telle que décrite aux Eqs. (17) et (18) et suite à une estimation des paramètres Ψ par descente de gradient sur la fonction objective $\mathcal{L}(\Psi)$ donnée à l'Eq. (19). De ces paramètres estimés, les plus importants sont évidemment les DoA des échos, $\tilde{\mathbf{\Omega}}_1 \approx [-121.2^\circ, 41.3^\circ]$ et $\tilde{\mathbf{\Omega}}_2 \approx [27.5^\circ, 57.5^\circ]$; soit des erreurs de 1.02° et 5.53° , respectivement.

Comme attendu, les amplitudes relatives sont bien estimées aussi : la deuxième réflexion est à $20 \log_{10} \left(\frac{|\tilde{h}_2|}{|\tilde{h}_1|} \right) \approx -6.07$ dB.

À noter que la descente de gradient est effectuée sur la grille Sloan-Womersley de taille $D_2 = 361$ mais que la Figure 5 montre la fonction de localisation reconstruite sur la grille "de visualisation" de taille $D_3 = 1369$.

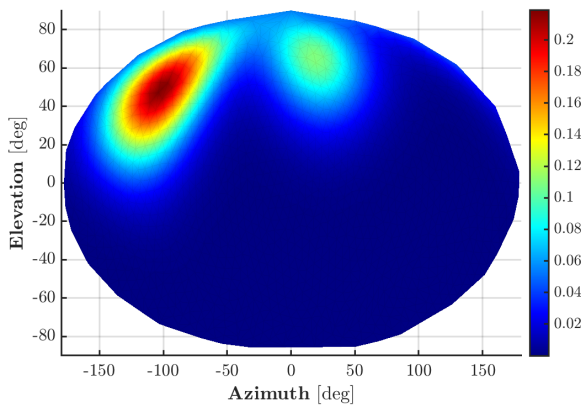


FIGURE 5 – Fonction de localisation reconstruite $\hat{\mathbf{m}}(\Psi)$, Eqs. (17) et (18), pour une trame contenant $P = 2$ échos placés à $\Omega_1 = [-120^\circ, 45^\circ]$ et $\Omega_2 = [22.5^\circ, 60^\circ]$, respectivement, suite à une descente de gradient sur la fonction objective $\mathcal{L}(\Psi)$ donnée à l'Eq. (19).

6 Conclusion

Cette communication a présenté les fondements d'une méthode de détection et de caractérisation des premières réflexions dans des réponses impulsionnelles spatiales de salles (SRIR) basée sur le formalisme de la fonction d'onde de Herglotz. Ce formalisme permet de décrire le champ mesuré par une antenne sphérique de microphones (SMA) comme une superposition d'ondes planes munies d'un noyau (*Herglotz kernel*) représentant la probabilité de la présence de l'onde sur la surface de la sphère.

La localisation des réflexions incidentes dans une trame temporelle donnée passe alors par une détection de pics sur une fonction de localisation construite à partir d'une estimation des noyaux de Herglotz. Ici, nous avons modélisé cette fonction de localisation par une superposition de gaussiennes radiales dont les paramètres sont estimés à travers une descente de gradient. Ces paramètres donnent à leur tour une estimation des directions d'arrivée (DoA) des échos incidents ainsi que de leurs amplitudes relatives.

Nous nous sommes limités ici à une étude sur des signaux synthétiques simulant l'arrivée de deux réflexions parfaitement blanches au sein d'une seule trame temporelle. La prochaine étape de ce travail consistera à étendre cette étude fondamentale en examinant notamment le comportement de la méthode face à de plus grands nombres d'échos incidents dans des configurations (répartition sur la sphère, amplitudes relatives) différentes.

Ensuite, il sera évidemment nécessaire d'appliquer la méthode dans le cadre d'une analyse trame-par-trame du segment précoce d'une SRIR, d'abord simulée (p. ex. par sources-images) et ensuite mesurée par SMA en conditions "réelles". Le résultat de ce travail renseignera enfin la pertinence de l'approche présentée à des fins de manipulation et de modification des caractéristiques

spatiales et énergétiques de SRIR.

Remerciements

Les auteurs voudraient remercier Wolfgang Kreuzer (*Acoustics Research Institute, Académie autrichienne des sciences, Vienne*) pour sa collaboration sur l'utilisation du formalisme de la fonction d'onde de Herglotz. Une partie de ce travail a été réalisé lors d'une thèse à l'École doctorale Informatique, traitement du signal, électronique de Paris (EDITE). Ce travail a aussi été financé en partie par le projet RASPUTIN (ANR-18-CE38-0004, <https://rasputin.lam.jussieu.fr>).

Références

- [1] E. G. Williams, *Fourier Acoustics : Sound Radiation and Nearfield Acoustical Holography*, Academic Press, Cambridge (1999).
- [2] B. Rafaely, Analysis and Design of Spherical Microphone Arrays, *IEEE Trans. Speech Audio Process.* **13** 1, 135-143 (2005).
- [3] B. Rafaely, Spatial Sampling and Beamforming for Spherical Microphone Arrays, *Proc. 2008 HSCMA*, Trento, Italy (2008).
- [4] J. Daniel, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*, Thèse de doctorat, Université Paris 6 (2001).
- [5] J. Daniel & S. Moreau, Further Study of Sound Field Coding with Higher Order Ambisonics, *Proc. 116th A.E.S. Conv.*, Berlin, Germany (2004).
- [6] Z. Li & R. Duraiswami, Flexible and Optimal Design of Spherical Microphone Arrays for Beamforming, *IEEE Trans. Audio Speech Lang. Process.* **15** 2, 702-714 (2007).
- [7] B. Rafaely, Plane-Wave Decomposition of the Sound Field on a Sphere by Spherical Convolution, *J. Acoust. Soc. Am.* **116** 4, 2149-2157 (2004).
- [8] T. D. Abhayapala & D. B. Ward, Theory and Design of High Order Sound Field Microphones Using Spherical Microphone Array, *Proc. 2002 IEEE I.C.A.S.S.P.*, Orlando, U.S.A. (2002).
- [9] M. Noisternig, F. Zotter, & B. F. G. Katz, Reconstructing Sound Source Directivity in Virtual Acoustic Environments, *Principles and Applications of Spatial Hearing*, World Scientific Publishing, Singapore (2011).
- [10] D. Colton & R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, Springer, New York (2013).
- [11] B. Jo, F. Zotter, & J.-W. Choi, Extended Vector-Based EB-ESPRIT Method, *IEEE/ACM Trans. Audio Speech Lang. Proc.* (2020).
- [12] I. H. Sloan & R. S. Womersley, Extremal Systems of Points and Numerical Integration on the Sphere, *Advances in Computational Mathematics* **21** 1-2, 107-125 (2004).
- [13] C. L. Lawson, C^1 Surface Interpolation for Scattered Data on a Sphere, *Rocky Mountain Journal of Mathematics*, **14** 1, 177-202, (1984).