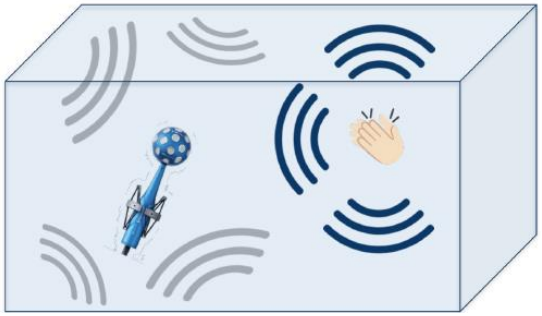
JOURNEES TECHNIQUES ACOUSTIQUE ET VIBRATIONS

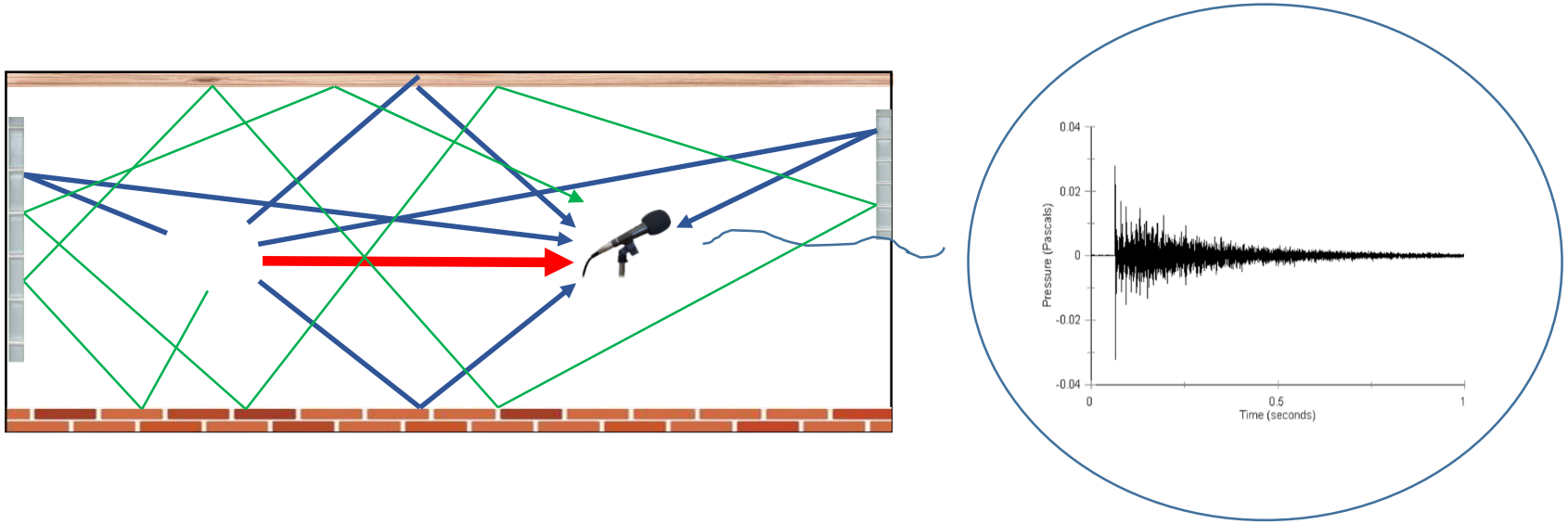
Détermination des propriétés acoustiques et géométriques d'une salle par méthodes inverses

Présentateur : Cédric Foy – Chargé de Recherche UMRAE

Bron le 12-13 mai 2022



Problématique



<u>Objectif</u> : extraction de ces paramètres d'entrée $\boldsymbol{X}$ influençant le champ sonore d'une salle (paramètre de sortie $\boldsymbol{Y}$)

La Réponse impulsionnelle d'une salle (RIR) contient toute les informations relatives aux paramètres d'entrée $\boldsymbol{X}$ qui ont agit sur le champ sonore (absorption des parois, dimensions de la salle, position des sources et des récepteurs…..)

Méthodologie

Comment extraire ces paramètres d'entrée X ?

<u>Principe</u> :

- Ensemble d'observations acoustiques issues d'un modèle théorique $[\mathbf{Y}(\boldsymbol{X})]$
- Ensemble d'observations acoustiques «réelles » de la salle $[\mathbf{Y}(\boldsymbol{X}?)\ ]$
- Post-traitement éventuel des observations en fonction des paramètres d'entrée $[\boldsymbol{X}]$ à atteindre
- Problème d'optimisation : On cherche alors les paramètres d'entrée $[\boldsymbol{X}]$ qui vont minimiser la différence "$[\mathbf{Y}([\boldsymbol{X}]\ ?)] - [\mathbf{Y}(\boldsymbol{X})]$" par des méthodes d'apprentissage

$$\underset{[\boldsymbol{X}]}{argmin} \quad ([\mathbf{Y}([\boldsymbol{X}]\ ?)\ ] - [\mathbf{Y}(\boldsymbol{X})])$$

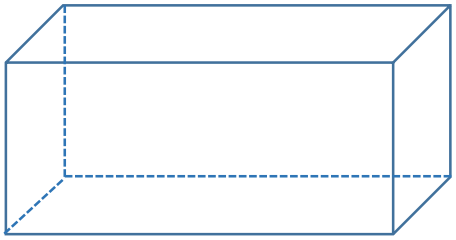
- Choix d'un algorithme de résolution
- $[\boldsymbol{X}] \equiv [\boldsymbol{X}\ ]$

CHOIX D'UN MODELE THEORIQUE

Modèle théorique

- Milieu fermé D + source étendue

$$\begin{cases} \left[\Delta + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right] p(\boldsymbol{r},t) = f(\boldsymbol{r},t) \quad \textit{équation volumique} \\ \frac{\partial p}{\partial n} + \frac{1}{c}\frac{\partial \beta(t)}{\partial t} * p(\boldsymbol{r},t) = 0 \ \textit{condition aux limites} \end{cases}$$



$\beta(t)$ admittance acoustique des parois

- Milieu fermé D + source impulsionnelle en temps et en espace

$$\begin{cases} \left[\Delta + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right] G(\boldsymbol{r},\boldsymbol{r}_{sce};t,t_{sce}) = -\delta(\boldsymbol{r}-\boldsymbol{r}_{sce})\delta(t-t_{sce}) \\ \frac{\partial p}{\partial n} + \frac{1}{c}\frac{\partial \beta}{\partial t} * G(\boldsymbol{r},\boldsymbol{r}_{sce};t,t_{sce}) = 0 \end{cases}$$

Fonction de Green

$$G(\boldsymbol{r},\boldsymbol{r}_{sce};t,t_{sce}) = \frac{1}{4\pi\|\boldsymbol{r}_{sce}-\boldsymbol{r}\|}\delta\left(\frac{\|\boldsymbol{r}-\boldsymbol{r}_{sce}\|}{c} - (t-t_{sce})\right)$$

- Du fait de la linéarité des équations, en décomposant la source étendue en une somme de sources élémentaires, la solution de l'équation volumique a la forme d'une intégrale :

$$\mathrm{p}(\boldsymbol{r},t) = \iint G(\boldsymbol{r},\boldsymbol{r}_{sce};t,t_{sce}) f(\boldsymbol{r}_{sce},t_{sce}) d\boldsymbol{r}_{sce}\, dt_{sce}$$

Modèle théorique

Intégration de la condition aux limites par la méthode des sources-images : passage d'un milieu fermé à une source à un milieu ouvert avec une infinité de sources (-images) dont les amplitudes associées intègrent les phénomènes aux parois

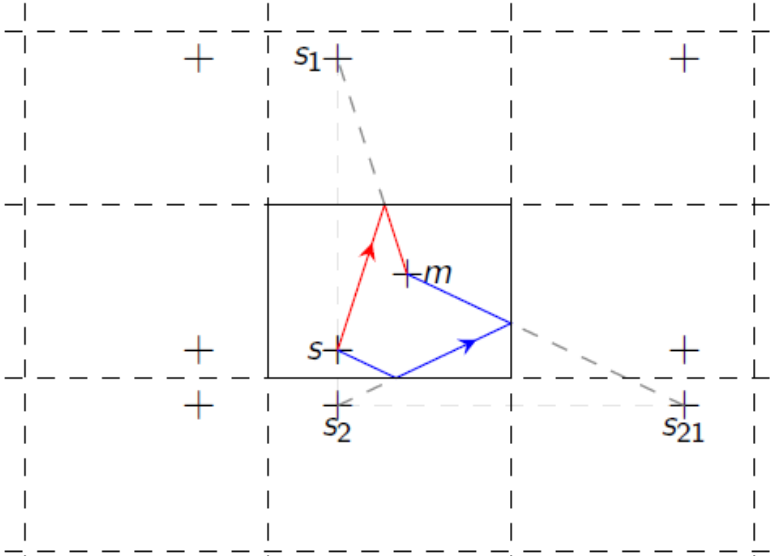


Image method for efficiently simulating small-room acoustics, Allen and Berkley, JASA 1979

"The image solution of a rectangular enclosure rapid/y approaches an exact solution of the wave equation as the walls of the room become rigid."

"If the room walls are not rigid.....we have continued to assume the approximate point image model even for non rigid walls."

$$f(\boldsymbol{r}_{sce}, t_{sce}) = \sum_{k=0}^{\infty} a_k(t)\delta(\boldsymbol{r}_{sce} - \boldsymbol{r}_k)\delta(t_{sce} - \tau)$$

avec $\tau = 0$ *(souces images synchrones)*

$k = 0$ *source réelle*

$k \neq 0$ *source − image numéro k*

$$p(\boldsymbol{r}, t) = \sum_{k=0}^{\infty} \frac{a_k(t)}{4\pi\|\boldsymbol{r} - \boldsymbol{r}_k\|}\delta(t - t_k)$$

$$t_k = \frac{\|r - \boldsymbol{r}_k\|}{c}$$

Temps d'arrivée de la source-image k *(TOA)*

lié à l'amplitude la source réelle

$$a_k(t) = 4\pi a_0(t) \prod_j R_j\left(\beta_j(t)\right) = 4\pi a_0(t)\boldsymbol{R}_k$$

lié aux parois rencontrées

$R_j(t)$ = coefficient de réflexion de la paroi j

Modèle théorique

- Filtrage $F(t)$ prenant en compte la chaîne de mesure

$$\mathrm{p_F}(\boldsymbol{r},t) = \left(p * F(t)\right) = \left[\sum_{k=0}^{\infty} \frac{a_k(t)}{4\pi\|\boldsymbol{r}-\boldsymbol{r_k}\|}\delta(t-t_k)\right] * F(t)$$

- Nombre limité de sources-images, pour une position de source s et de microphone m

$$\mathrm{p}_{s,m}(\boldsymbol{r},t) \approx \left[\sum_{k=0}^{K} \frac{a_{s,m,k}(t)}{4\pi\|\boldsymbol{r}-\boldsymbol{r_k}\|}\delta\left(t-t_{s,m,k}\right)\right] * \mathrm{F(t)}\ (+\mathrm{e(t)}) \equiv RIR$$

réponse impulsionnelle théorique de salle

$$a_{s,m,k}(t) = \frac{4\pi a_0(t)}{c t_{s,m,k}} \prod_j R_j\left(\beta_j(t)\right) = a_0(t)\boldsymbol{R}_{s,m,k}(t)$$

$R_j(t)$ = coefficient de réflexion de la paroi j

Paramètre de sortie

$$\mathrm{p}_{s,m}(\boldsymbol{r},t) \equiv Y_j$$

$$a_0(t)?\,, t_{s,m,k}?\,, R_j(t)?\,, \boldsymbol{r_k}? \equiv X_j$$

Paramètres d'entrée

(Stéphane :) recherche des coefficients énergétiques d'absorption en fonction de la fréquence $\boldsymbol{\alpha_j(f) = (1 - R_j^2(f))}$

(Tom :) recherche des positions des sources-images $\boldsymbol{r_k}$ (salles rectangulaires, sources-images visibles)

X : coefficient énergétique d'absorption des parois en fonction de la fréquence

Stéphane Dilungana
Doctorant

Sylvain Faisan
Maître de conférences (HDR)
Université de Strasbourg - Icube

Antoine Deleforge
Chargé de Recherche
Inria Nancy - Multispeech

CFA 2022 : Joint estimation of surface absorption profiles from room impulse responses

Merci Stéphane pour les figures de cette présentation!

Pré-traitement

$$p_{s,m}(\boldsymbol{r},t) = \sum_{k=0}^{K} \frac{a_{s,m,k}(t)}{4\pi\|\boldsymbol{r}-\boldsymbol{r_k}\|}\delta(t-t_{s,m,k}) * F(t)$$

Décomposition du modèle

$$p_{s,m}(\boldsymbol{r},t) = \sum_{k=0}^{K} h_{s,m,k}.\left(v_{s,m,k}(t) * F(t)\right)$$

$$h_k = \frac{1}{c\tau_{s,m,k}}$$

$$v_{s,m,k}(t) = a_0(t)\boldsymbol{R}_k(t)\delta(t-t_{s,m,k})$$

$$F(t) = sinc(t-t_k)$$

$$\left|\mathrm{STFT}(p_{s,m}(\boldsymbol{r},t))\right|^2$$

$$P(\tau,f) = \int p(t)w(t-\tau)e^{2\pi i f t}\,dt$$

$$P_{s,m}(\boldsymbol{r},t,f) = \sum_{k=0}^{K} H_{s,m,k}.V_k(t,f)$$

$$H_{s,m,k} = \left(\frac{1}{c\tau_{s,m,k}}\right)^2$$

Dépend de la source, du microphone et de la source-image k

$$V_k(t,f) = \left|STFT\left(v_{s,m,k} * F\right)\right|^2 = A_0^2\,(t,f)\boldsymbol{R}_k^2(t,f)$$

Dépend de la source-image k

Inconnues $\alpha_j(f) = (1 - R_j^2(f))$

Pré-traitement

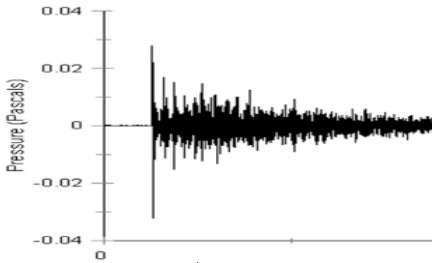
- Création des observations $\mathbf{Y}(\boldsymbol{X})$ - S sources réelles s et M microphones m (récepteurs)

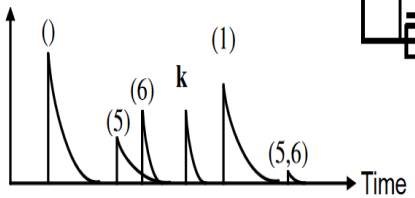
1 source s
1 récepteur m

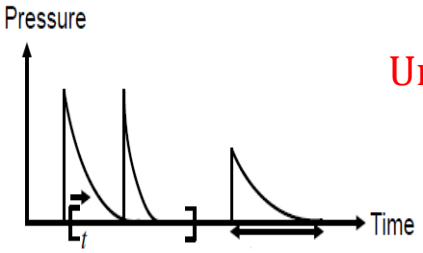
$$[P_{1,1}]_{[T,F]} \Rightarrow \underbrace{[[P_{1,1}] \quad \dots \quad [P_{s,m}] \quad \dots \quad [P_{S,M}]]}_{[T.M.S,F]}$$

S sources s, M microphones m

$$[H][V](f) = \begin{bmatrix} [H_{1,1,0}] & \dots & [H_{m,s,k}] & \dots & [H_{M,S,K}] \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ [H_{m,s,0}] & \dots & [H_{m,s,k}] & \dots & [H_{M,S,K}] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ [H_{1,1,0}] & \dots & [H_{m,s,k}] & \dots & [H_{M,S,K}] \end{bmatrix}_{[T.M.S,K]} \begin{bmatrix} [V_{1,1,0}] & \dots & [V_{m,s,0}] & \dots & [V_{M,S,0}] \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ [V_{m,s,k}] & \dots & [V_{m,s,k}] & \dots & [V_{M,S,k}] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ [V_{1,1,K}] & \dots & [V_{m,s,K}] & \dots & [V_{M,S,K}] \end{bmatrix}_{[K,F]}$$







Un grand nombre d'observations inexploitables

$$P(\tau, f) = \int p(t) w(t-\tau) e^{2\pi i f t} dt$$

$w(t)$ fenêtre glissante de largeur temporelle L_{STFT}

L_k : largeur temporelle d'une réflexion $v_k(t)$

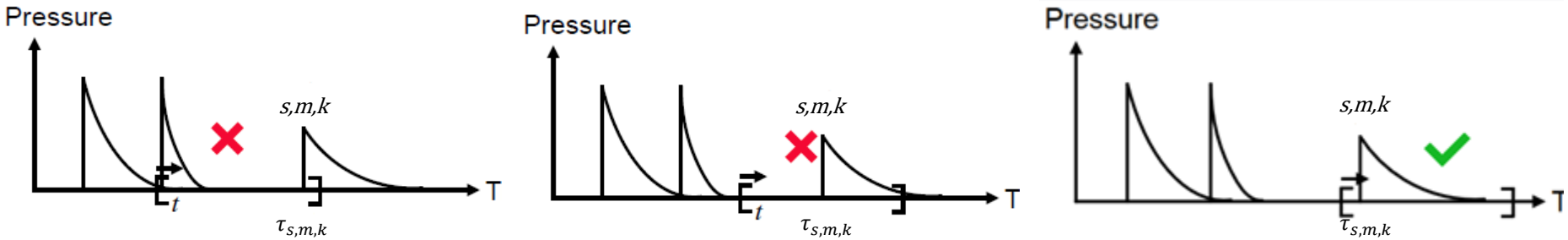
Pré-traitement

L_k : largeur temporelle d'une réflexion $v_k(t)$

L_{STFT} : largeur temporelle de la fenêtre de la STFT

- Proposition 1 : sélection des fenêtres temporelles

$L_{STFT} > L_k$



$W^- = [t : t + L_{DFT} - L_k]$ $\quad$ $W^+ = [t - L_k : t + L_{DFT} - 1]$

$$\mathcal{G}_k^* = \{(s,m,t) | \tau_{s,m,k} \in W^-, \tau_{s,m,k'} \notin W^+\}$$

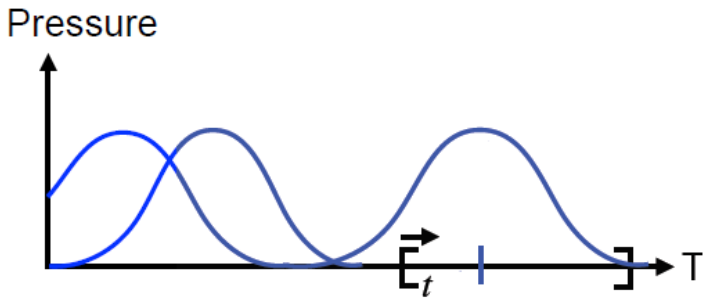
Ensemble correspondant aux fenêtres temporelles contenant théoriquement uniquement une seule source image

Cet ensemble est défini à partir des TOA théoriques des sources-images supposés connus

En pratique, il y a toujours des incertitudes sur la connaissance de la salle et sur les TOA. De ce fait, ces fenêtres peuvent ne pas inclure uniquement la réflexion k d'intérêt.

Pré-traitement

- Proposition 2 : aspect probabiliste



Pour tenir compte du fait que les positions des sources, microphones et surfaces sont connues à une incertitude près σ_k,

$$p(\tau_{k\ réel}) = N\left(\tau_{k\ réel}, \tau_{k\ théo}, \sigma_k^2\right)$$

$$\sigma_k = (q(k)+2)\ \sigma_{geo}$$

$q(k)$ ordre de la source-image k

Finalement toutes les fenêtres sélectionnées dans l'ensemble précédent sont associées à une probabilité de présence uniquement de la source-image k

$$\pi_{s,m,t,k} = \left\{(s,m,t) \middle| p(\tau_{s,m,k}) \in W^-, p(\tau_{s,m,k'}) \notin W^+\right\}$$

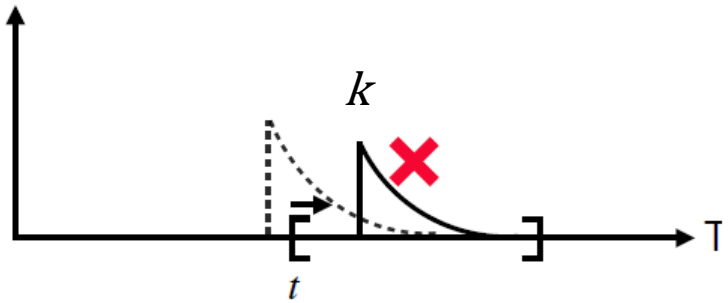
Pour chaque source-image k, on retient uniquement les fenêtres ayant la plus forte probabilité

$$\boxed{g_k^{top} = \{(s,m,t) | argmax \quad \pi_{s,m,t}\}}$$

- 1 seule fenêtre pour chaque réflexion k
- Approche déterministe du fait du argmax
- Malheureusement g_k^{top} peut être très petit par rapport à g_k^*

Pré-traitement

- Proposition 3 : approche robuste RANSAC



Pour g_k^* et g_k^{top}

Approche moins fiable à mesure pour les sources images d'ordre élevées, car plus forte densité des ces

certaines fenêtres sélectionnées peuvent contenir des informations erronées en raison d'interférences avec réflexions d'ordre supérieur plus élevé (non pris en compte mais existant) ou à cause du bruit

g_k^{ransac}

Développement d'une méthode robuste inspirée de l'algorithme RANSAC

*« **RANSAC**, abréviation pour **RANdom SAmple Consensus**, est une méthode itérative utilisée lorsque l'ensemble de données observées peut contenir des valeurs aberrantes (outliers). Il s'agit d'un algorithme non-déterministe dans le sens où il produit un résultat correct avec une certaine probabilité, celle-ci augmentant à mesure que le nombre d'itérations est grand. L'hypothèse de base est que les données sont constituées d' inliers et d'outliers. De plus, les données peuvent être soumises au bruit. »*

Pré-traitement

- Proposition 3 : approche itérative robuste RANSAC

$$\pi_{s,m,t,k} = \left\{(s,m,t) \middle| p(\tau_{s,m,k}) \in W^-, p(\tau_{s,m,k'}) \notin W^+\right\} \quad \text{(proposition 2)}$$

Itération n : On tire un jeu de 7 (s,m,t)

$$[A_0^2] = \frac{[Y]}{[H_k][1]}$$ Calcul du champ direct

$$[\boldsymbol{R}_k^2] = \frac{[Y]}{[H_k][A_0^2]}$$ Pour sources-images k d'ordre 1 (i.e. 6 sources-images)

$$\hat{g}_k^{ransac} = \left\{(s,m,t) \middle| \left|\frac{[Y] - [H_k][A_0^2][\boldsymbol{R}_k^2]}{[H_k][A_0^2][\boldsymbol{R}_k^2]}\right| < 0{,}1\right\}$$ Critère d'acceptabilité

On compte le nombre d'inliers dans $\hat{g}_k^{ransac}$

Itération n+1 : On tire un nouveau jeu de 7 (s,m,t)

.....

.....

On retient $\hat{g}_k^{ransac}$ associé à l'itération ayant donnée le plus d'inliers. $\boxed{g_k^{ransac}}$

Problème d'optimisation

Identique pour les 3 ensembles $\mathcal{g}_k^*$, $\mathcal{g}_k^{top}$ et $\mathcal{g}_k^{ransac}$

$$\underset{a_0(f) et\ \boldsymbol{R}_j(f)}{argmin} \quad ([\boldsymbol{Y}(\boldsymbol{f})] - [\boldsymbol{H}][\boldsymbol{V}](\boldsymbol{f}))$$

$$\Downarrow$$

$$\alpha_j(f) = (1 - R_j^2(f))$$

$$\Downarrow$$

$$\alpha_j(f) \equiv \alpha_j(f)$$

Géométrie , position sources et microphones connus pour estimer les TOA

Algorithme de résolution

Descente de gradient (fonction Fmincon de Matlab)

Ensemble d'observations

- Données simulées
- 500 salles de dimensions comprises entre : $3\text{m} < L_x < 10\text{m}$, $3\text{m} < L_y < 10\text{m}$, $2\text{m} < L_z < 5m$
- 3 positions de source et de microphone choisies aléatoirement, *i.e.* 9 RIR (contrainte : à plus d' 1 m des parois)
- Absorption des parois issues de données réalistes
- $\mathbf{Y}(\boldsymbol{X})$: données simulées (Roomsim) $\mathbf{Y}(\boldsymbol{X})$ + Bruit gaussien (PSNR)
- $L_{DFT} = 2ms$, $L_k = 0{,}5ms$, $\sigma_{geo} = 2\text{cm ou } 5\text{cm}$
- Durée des RIR : 250 ms (f_e=16kHz), sources-images jusqu'à l'ordre q=20
- f < 400 Hz sont exclues (dualité temps – fréquences)
- Algo Ransac : 2000 itérations

Résultats

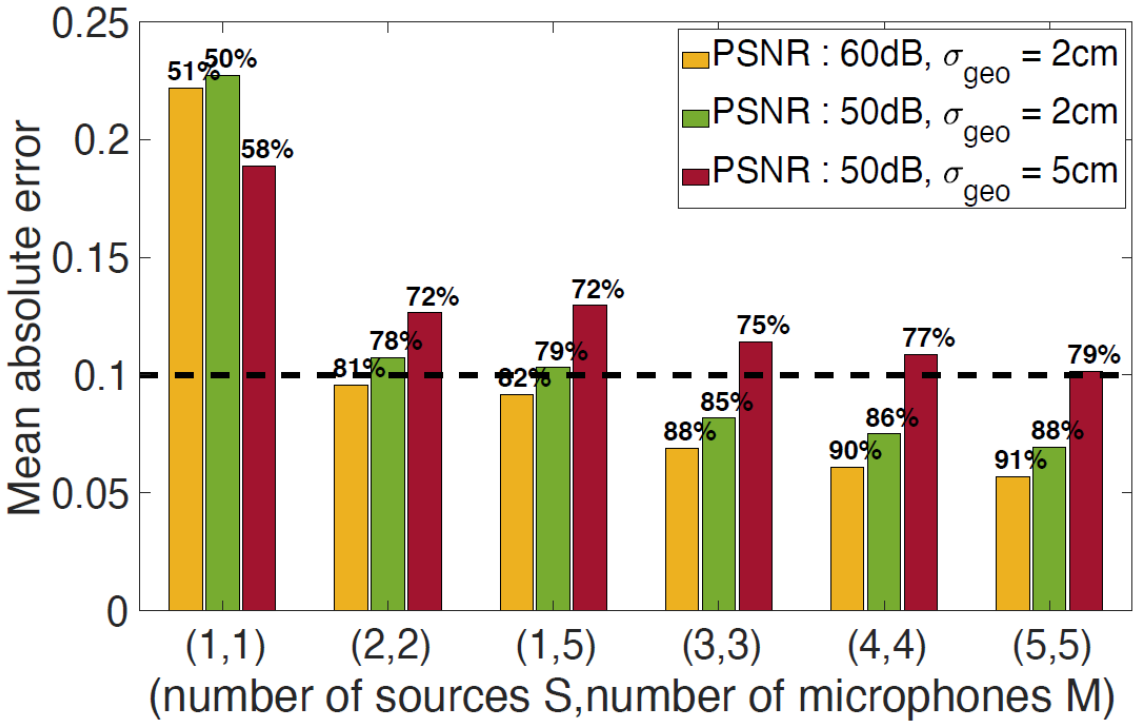
	Sources-images ordre 0&1		Sources-images ordres 0&1&2	
	MAE	CE (%)	MAE	CE (%)
g_k^*	0.108	74.7	0.132	69.6
g_k^{top}	0.128	71.7	0.119	69.7
g_k^{ransac}	0.088	82.9	**0.082**	**84.6**

MAE : erreur absolue moyenne (mean absolute error)
CE : estimation correcte

- l'erreur absolue est acceptable.
- g_k^* (cas idéal) la prise en compte des ordres 2 tend à pénaliser le résultat.
- g_k^{top} (cas plus réaliste) la prise en compte d'une incertitude géométrique dégrade que légèrement l'erreur absolue. La prise en compte des ordres 2 diminue cette erreur (plus cohérent).
- g_k^{ransac} l'algorithme RANSAC diminue nettement l'erreur absolue, d'autant plus si les ordres 2 sont considérés, avec un fort taux d'estimations correctes.

Résultats

$g_{k,f}^{ransac}$ et sources-images d'ordres 0&1&2



- Influence du bruit : l'erreur augmente légèrement avec le bruit.
- Influence de l'incertitude géométrique : sauf pour le cas (1,1), l'incertitude géométrique engendre une augmentation de l'erreur
- Influence du nombre de couples (sources, microphones) : cela tend à faire chuter l'erreur.

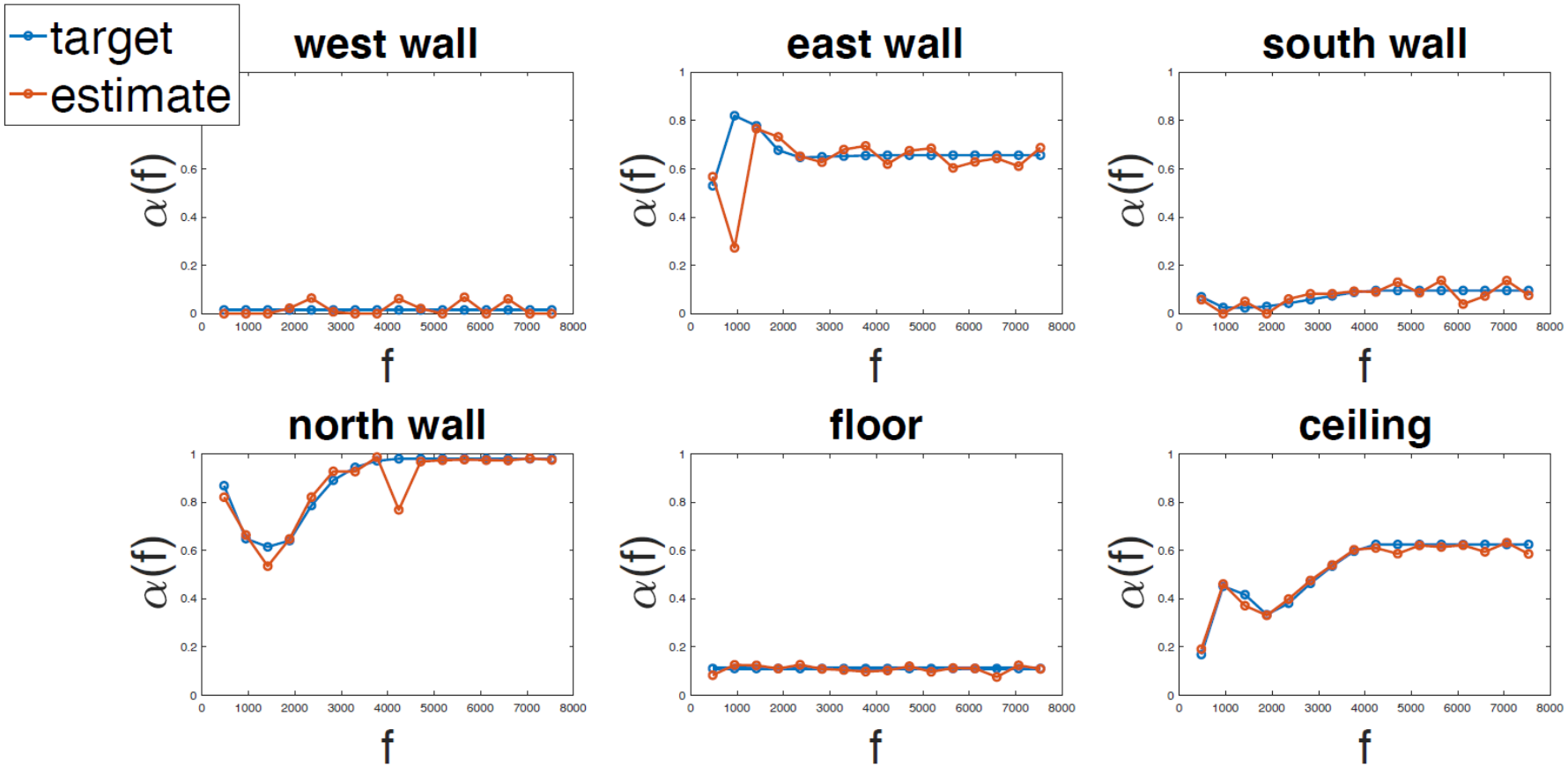
Résultats

$g_{k,f}^{ransac}$

$3 * 3\ RIR$

Sources-images d'ordres 0&1&2

$\sigma_{geo} = 2\text{cm}, \text{PSNR} = 50\text{dB}$



X : position des sources-images

Tom Sprunck
Doctorant

Yannick Privat
Professeur des Universités (HDR)
Université de Strasbourg - IRMA

Antoine Deleforge
Chargé de Recherche
Inria Nancy - Multispeech

CFA 2022 : Can one hear the shape of a room ?

Merci Tom pour les figures de cette présentation!

Pré-traitement

$$p_{s,m}(\boldsymbol{r},t) = \sum_{k=0}^{K} \frac{a_{s,m,k}}{4\pi\|\boldsymbol{r}-\boldsymbol{r_k}\|} F(t - t_{s,m,k})$$

Hypothèse : Amplitude indépendante en temps

Décomposition du modèle

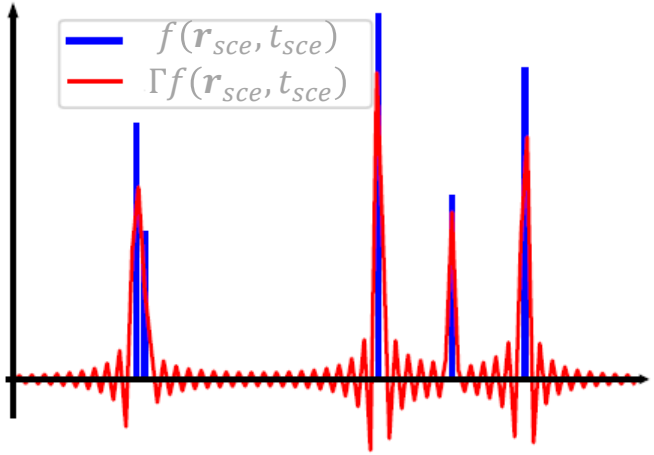
$$p_{s,m}(\boldsymbol{r},t) = \boldsymbol{\Gamma}\ f(\boldsymbol{r}_{sce}, t_{sce}) = \sum_{k=0}^{K} a_{s,m,k}\ \gamma(\boldsymbol{r_k},\boldsymbol{r},\boldsymbol{t})$$

$$f(\boldsymbol{r}_{sce}, t_{sce} = 0) = \sum_{k=0}^{\infty} a_{s,m,k}\delta(\boldsymbol{r}_{sce} - \boldsymbol{r_k})$$

$$F(t) = sinc(t - t_{s,mk})$$

$\boldsymbol{\Gamma}$ *opérateur linéaire avec le noyau* $\gamma(\boldsymbol{r_k},\boldsymbol{r},\boldsymbol{t}) = \dfrac{F(t-t_k)}{4\pi\|\boldsymbol{r}-\boldsymbol{r_k}\|}$

Dépend de la source-image k



La relation entre le terme source et le terme de réception est donnée par un opérateur linéaire Γ.
γ est une fonction non linéaire, non-convexe.

Problème d'optimisation

Blasso
(pour chaque simulation)

Ce problème inverse est un problème de super résolution où le but est de retrouver les positions continues des S-I à partir de mesures discrètes.

$$argmin \left[\underbrace{\frac{1}{2}\left\| \boldsymbol{Y} - \sum_{k=0}^{K} a_{s,m,k}\gamma(\boldsymbol{r_k}, \boldsymbol{r}, \boldsymbol{t}) \right\|^2}_{\text{Cohérence des données (moindre carré)}} + \underbrace{\boldsymbol{\lambda}\left\| \sum_{k=0}^{\infty} a_{s,m,k}\delta(\boldsymbol{r}_{sce} - \boldsymbol{r_k}) \right\|_{\boldsymbol{TV}}}_{\text{Terme de régularisation (gestion du bruit)}} \right]$$

$$a_{s,m,k}, \boldsymbol{r_k}$$

$$\boldsymbol{r_k} \equiv r_k$$

Pas de connaissance de la géométrie de la salle et de la position de la source-réelle, uniquement de la position des micros les uns par rapport aux autres.

Algorithme de résolution

Algorithme itératif de résolution : Sliding-Franck-Wolfe (SFW)

Sliding Franck Wolfe

Initialisation

Itération (i)

$$\Gamma f(\boldsymbol{r}_{sce}, t_{sce})^{(i)} = \sum_{k=0}^{K^{(i)}} a_{s,m,k}^{(i)} \, \gamma\left(\boldsymbol{r}_k^{(i)}, \boldsymbol{r}, \boldsymbol{t}\right)$$

Opérateur conjugué Γ^* tel que $\Gamma^* \boldsymbol{Y} = \sum_{j=1}^{J} \gamma\left(\boldsymbol{r}_k^{(i)}, \boldsymbol{r}, t_j\right) \boldsymbol{Y}(t_j)$

Calcul de $\eta^{(i)} = \frac{1}{\lambda} \Gamma^*\left[\boldsymbol{Y} - \Gamma f(\boldsymbol{r}_{sce}, t_{sce})^{(i)}\right]$

if $\left|\eta^{(i)}\right| < 1$: la solution est $a_{s,m,k}^{(i)}$ *et* $\boldsymbol{r}_k{}^{(i)}$

Else : $\widehat{\boldsymbol{r}_k}^{(i)} = \underset{\boldsymbol{r}_k{}^{(i)}}{\mathbf{argmax}} \left|\eta^{(i)}\right|$

I Position provisoire la source-image suivante

$$\boldsymbol{r}_k{}^{\left(i+\frac{1}{2}\right)} = \left[\boldsymbol{r}_k{}^{(i)}, \widehat{\boldsymbol{r}_k}^{(i)}\right]$$

$$\Gamma f(\boldsymbol{r}_{sce}, t_{sce})^{\left(i+\frac{1}{2}\right)} = \sum_{k=0}^{K^{(i)}+1} a_{s,m,k}^{\left(i+\frac{1}{2}\right)} \, \gamma\left(\boldsymbol{r}_k^{\left(i+\frac{1}{2}\right)}, \boldsymbol{r}, t\right)$$

$$\widehat{a_{s,m,k}}^{\left(i+\frac{1}{2}\right)} = \underset{a_{s,m,k}^{\left(i+\frac{1}{2}\right)}}{\mathbf{argmin}} \frac{1}{2}\left\|\boldsymbol{Y} - \Gamma f(\boldsymbol{r}_{sce}, t_{sce})^{\left(i+\frac{1}{2}\right)}\right\|^2 + \lambda \left\|a_{s,m,k}^{\left(i+\frac{1}{2}\right)}\right\|$$

II Amplitudes provisoires des source-images

$$\boldsymbol{r}_k{}^{(i+1)} = \left[\boldsymbol{r}_k{}^{(i)}, \widehat{\boldsymbol{r}_k}^{(i)}\right]$$

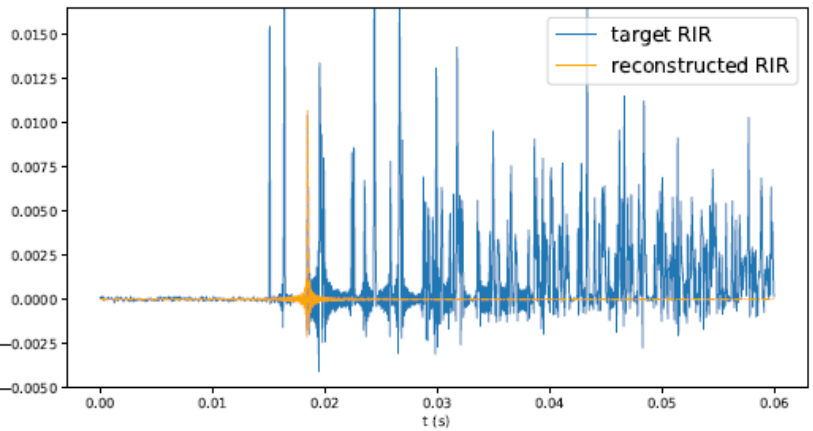
$$a_{s,m,k}^{(i+1)} = \left[a_{s,m,k}^{(i)}, \widehat{a_{s,m,k}}^{\left(i+\frac{1}{2}\right)}\right]$$

$$\widehat{a_{s,m,k}}^{(i+1)}, \widehat{\boldsymbol{r}_k}^{(i+1)} = \underset{\boldsymbol{r}_k{}^{(i+1)}, a_{s,m,k}^{(i+1)}}{\mathbf{argmin}} \frac{1}{2}\left\|\boldsymbol{Y} - \Gamma f(\boldsymbol{r}_{sce}, t_{sce})^{(i+1)}\right\|^2 + \lambda \left\|a_{s,m,k}^{(i+1)}\right\|$$

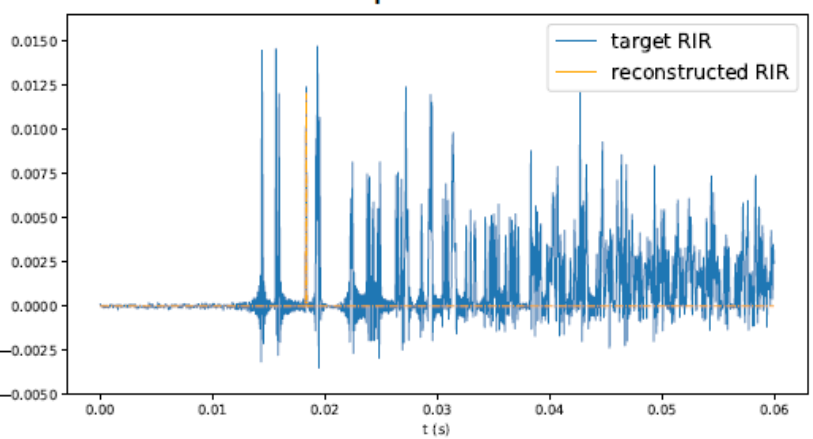
II Amplitudes et positions provisoires des source-images

Sliding-Frank-Wolfe [3], iteration 1

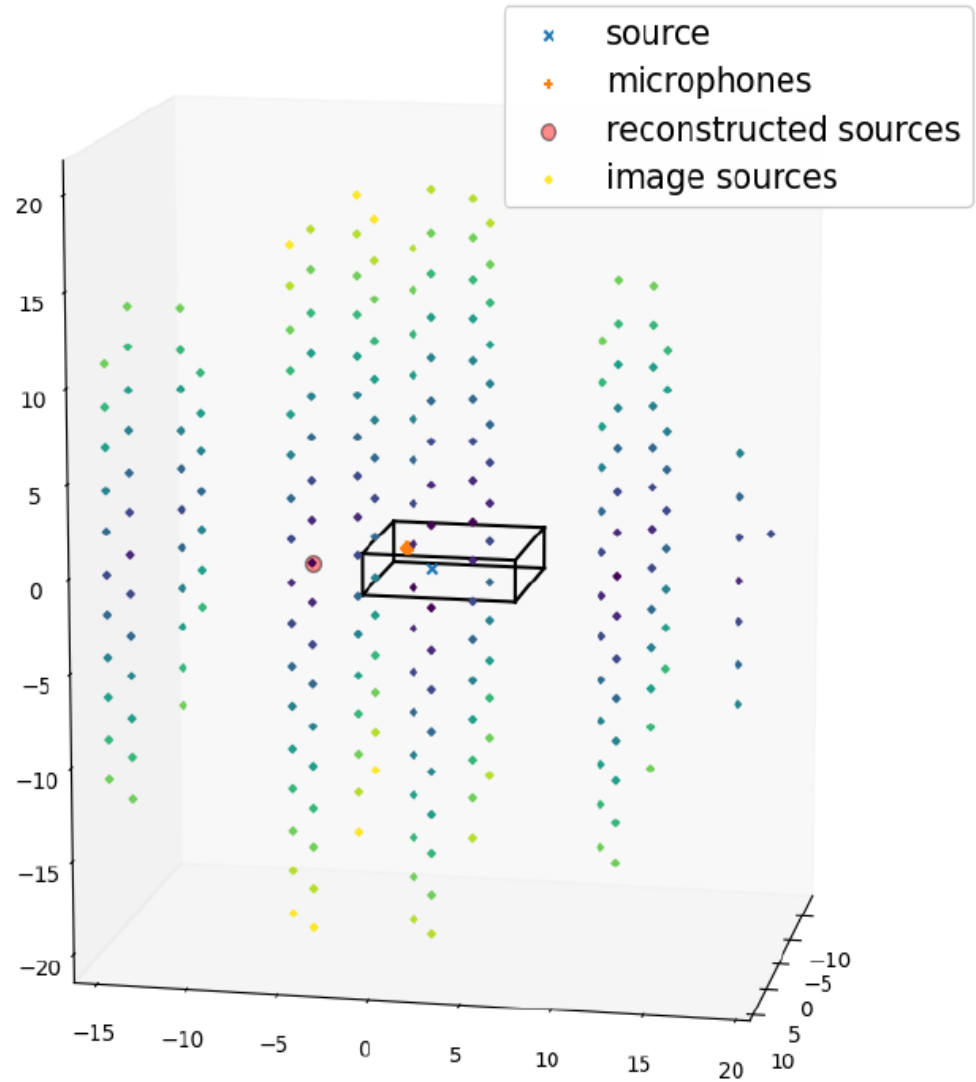
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

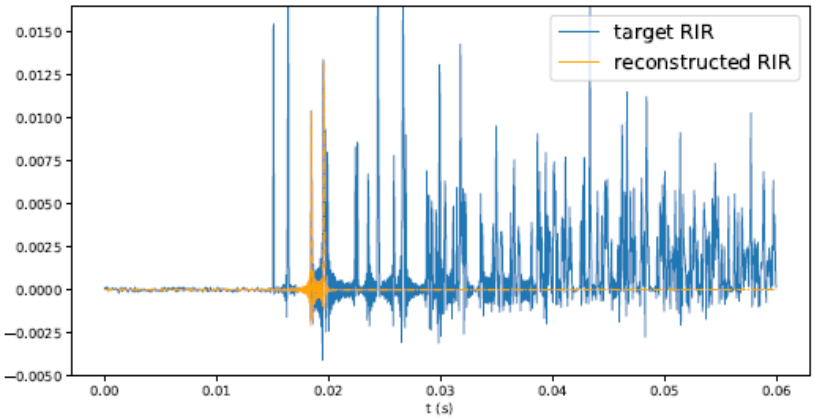


Microphone 2

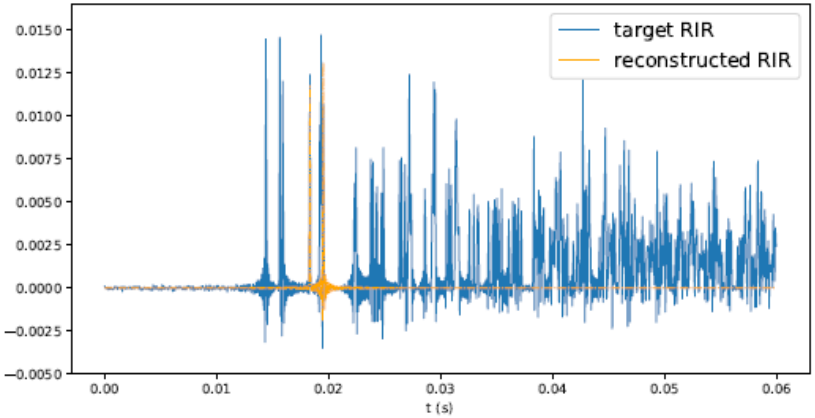


Sliding-Frank-Wolfe [3], iteration 2

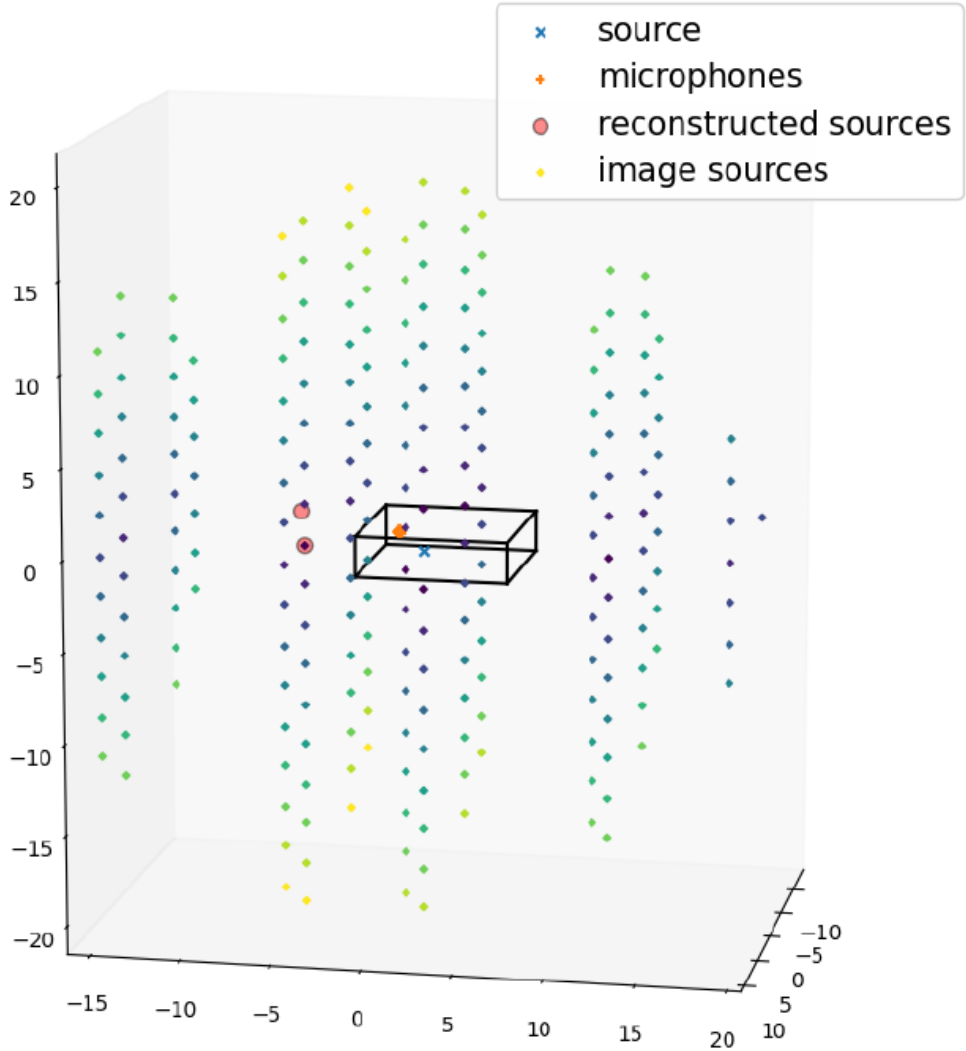
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

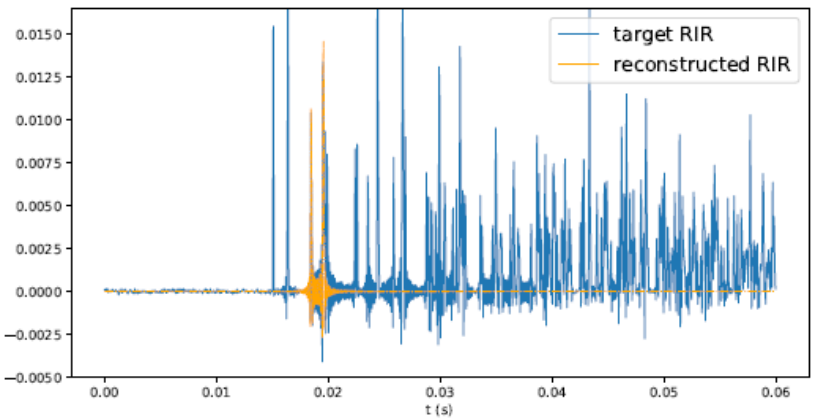


Microphone 2

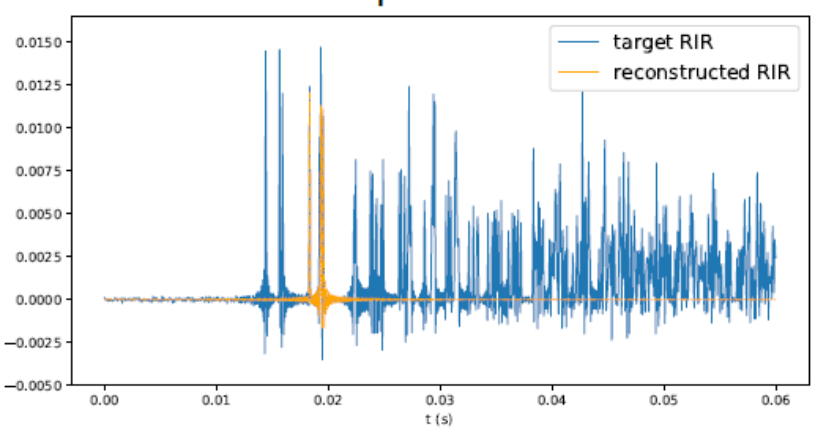


Sliding-Frank-Wolfe [3], iteration 3

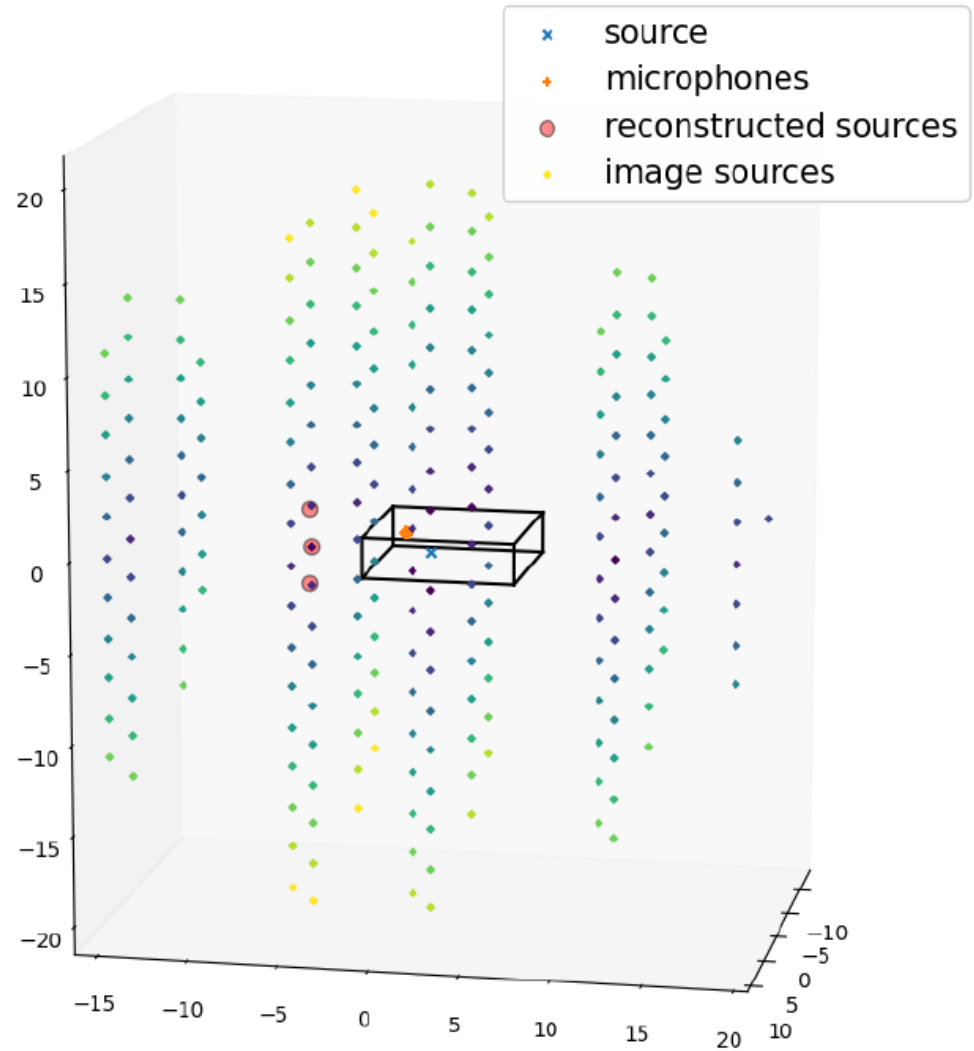
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

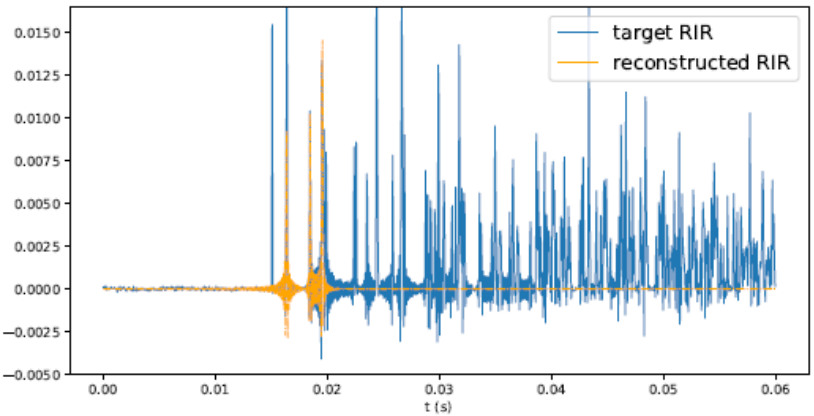


Microphone 2

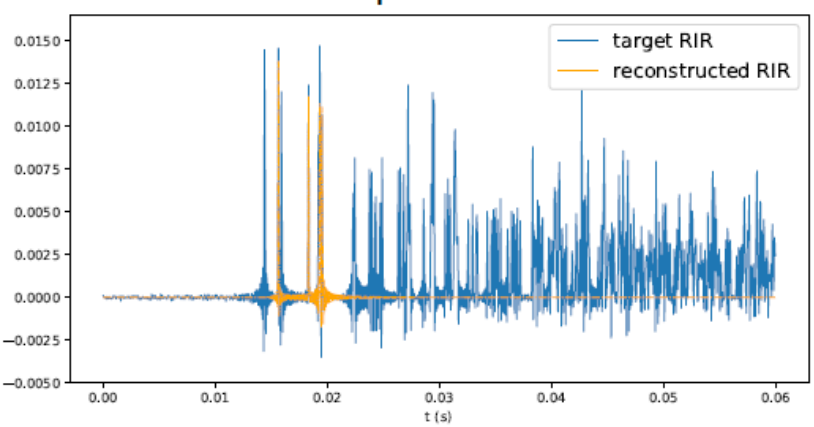


Sliding-Frank-Wolfe [3], iteration 4

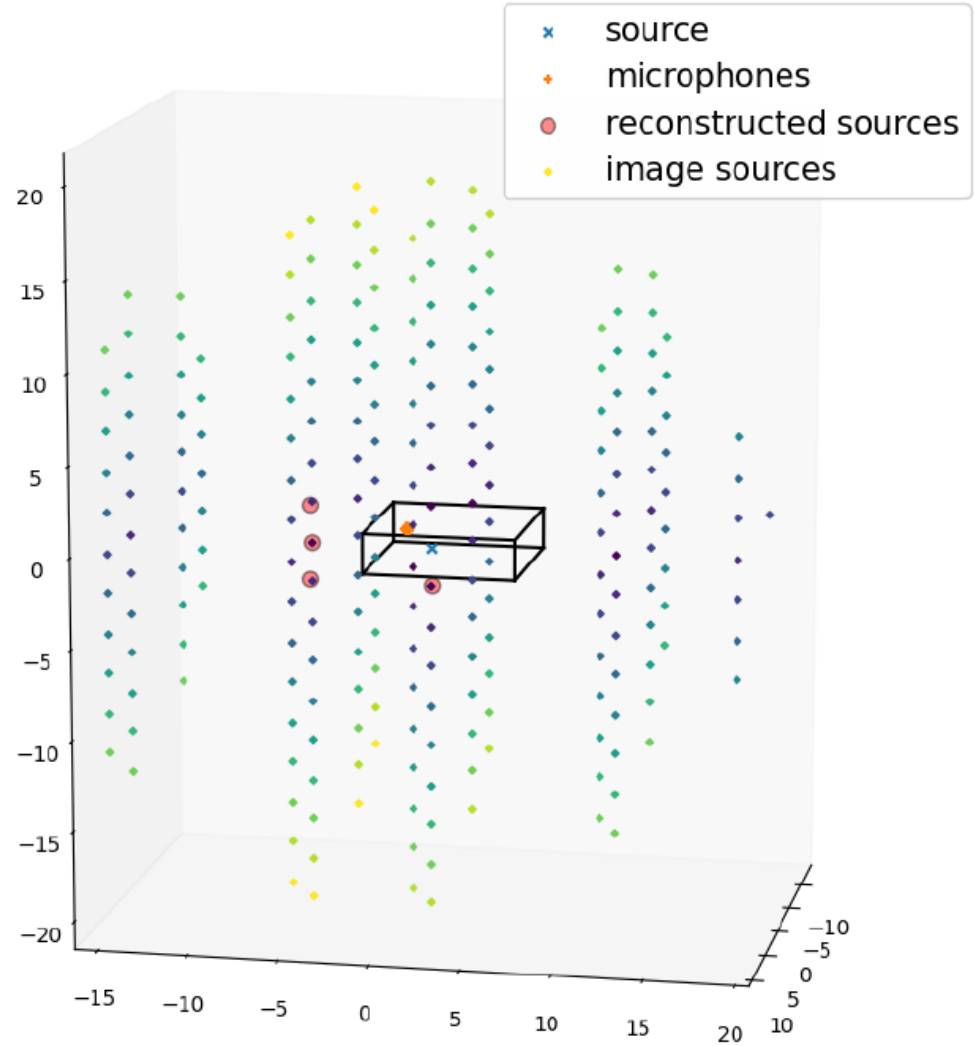
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

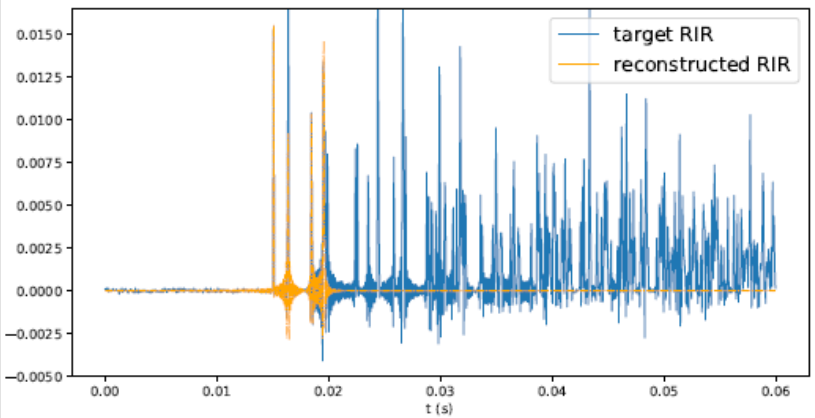


Microphone 2

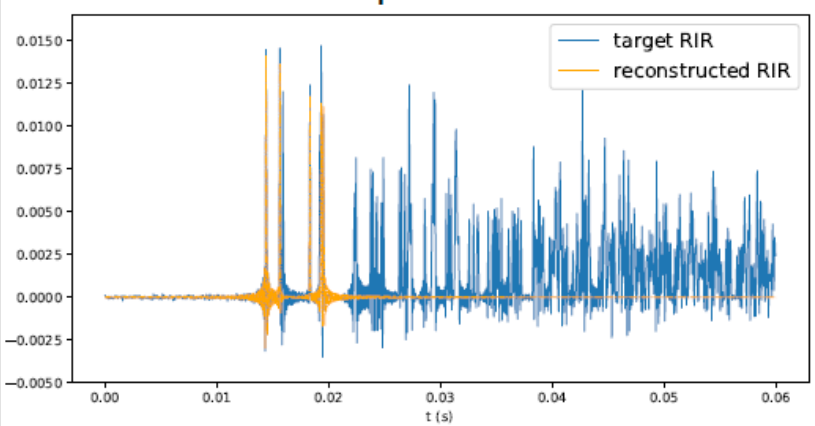


Sliding-Frank-Wolfe [3], iteration 5

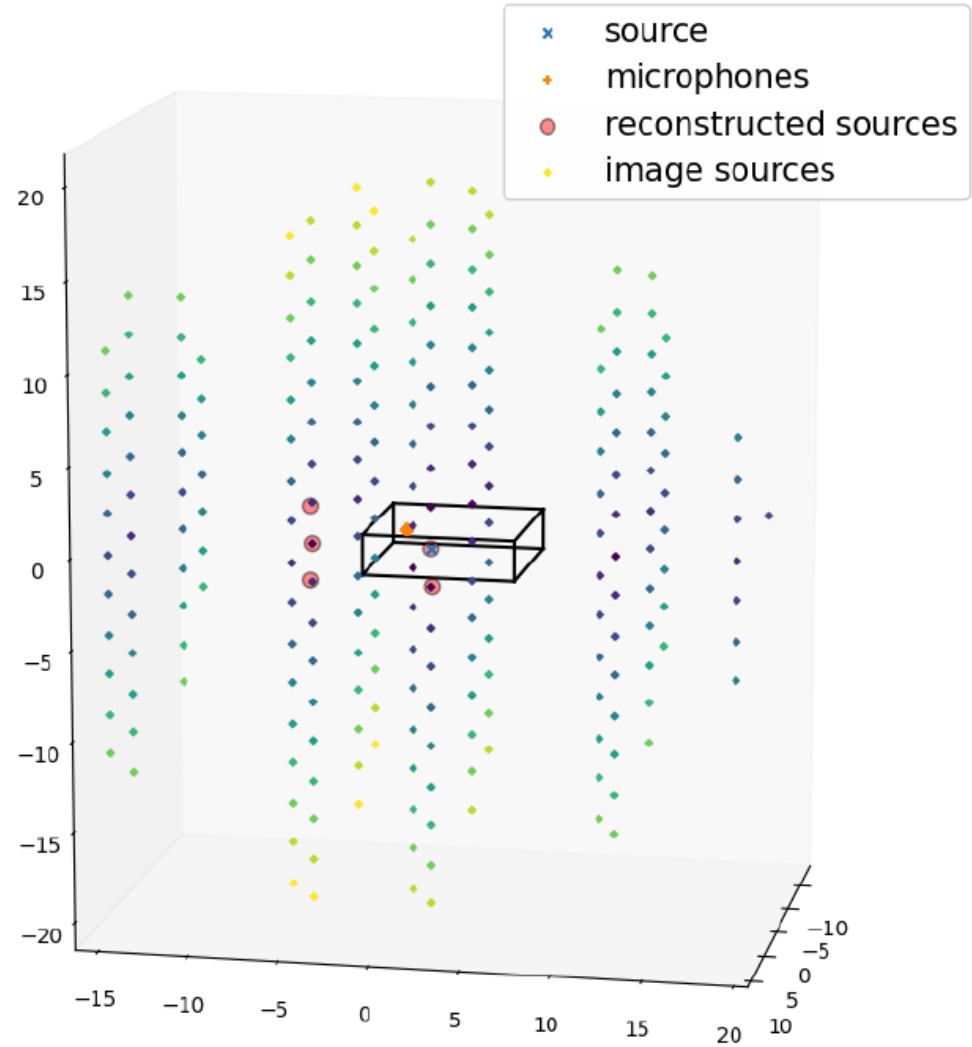
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

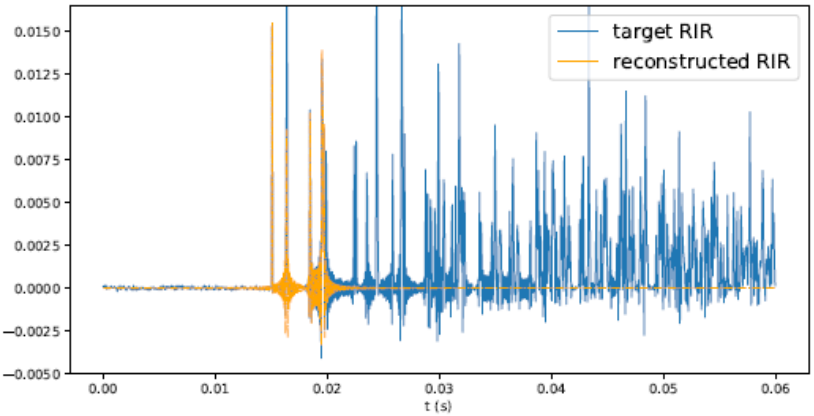


Microphone 2

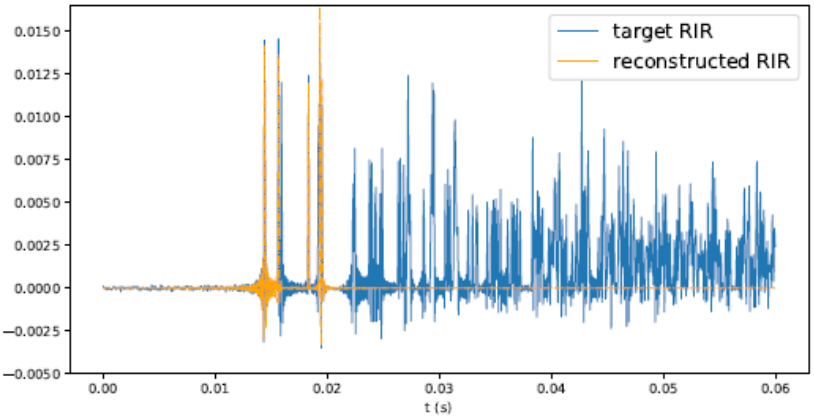


Sliding-Frank-Wolfe [3], iteration 6

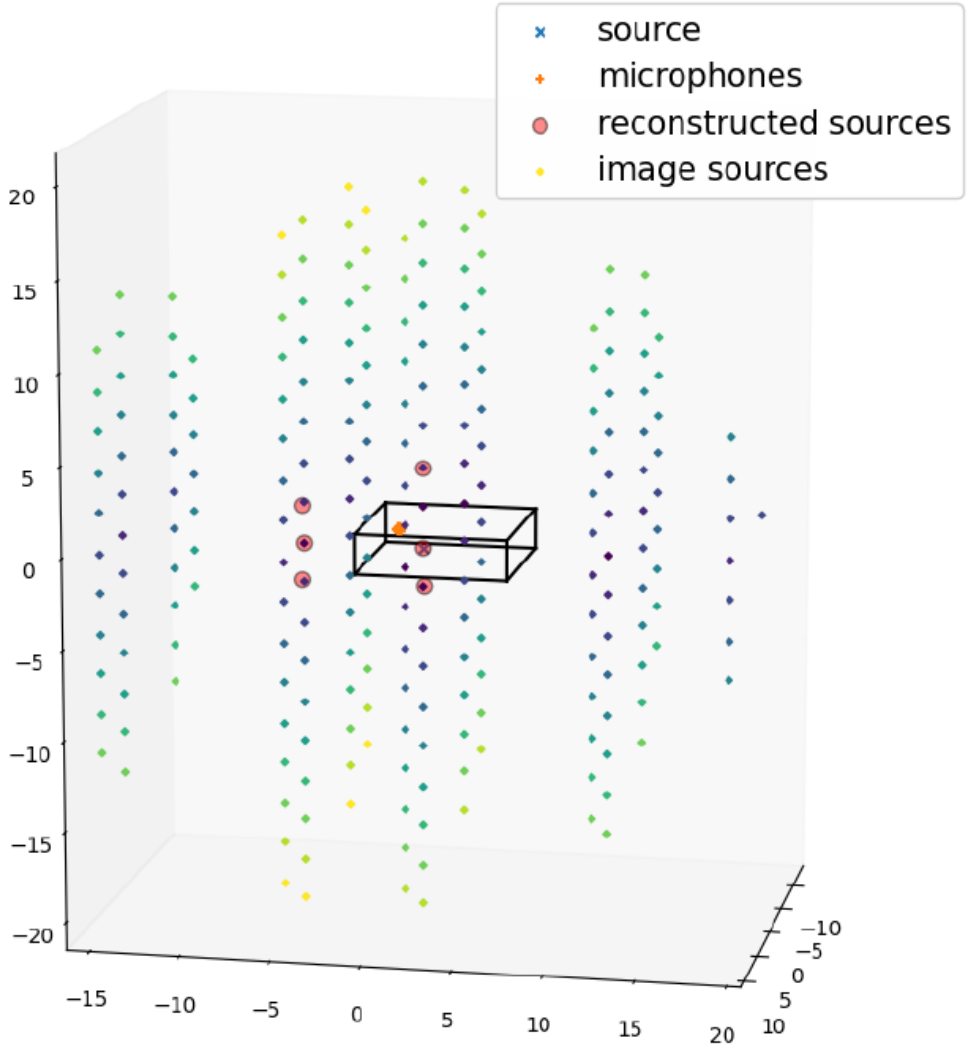
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

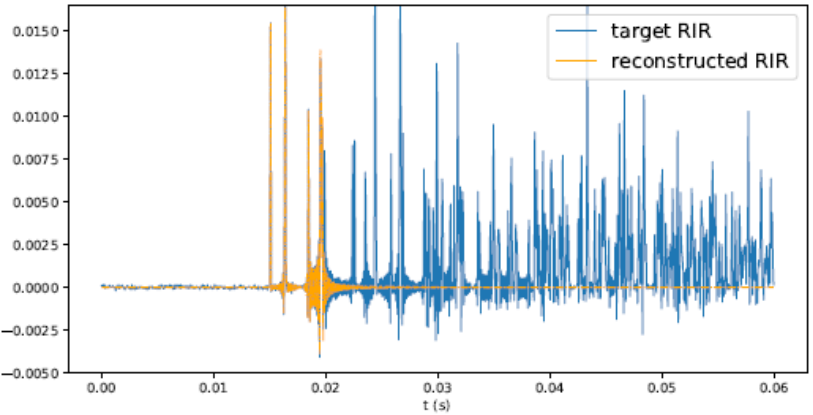


Microphone 2

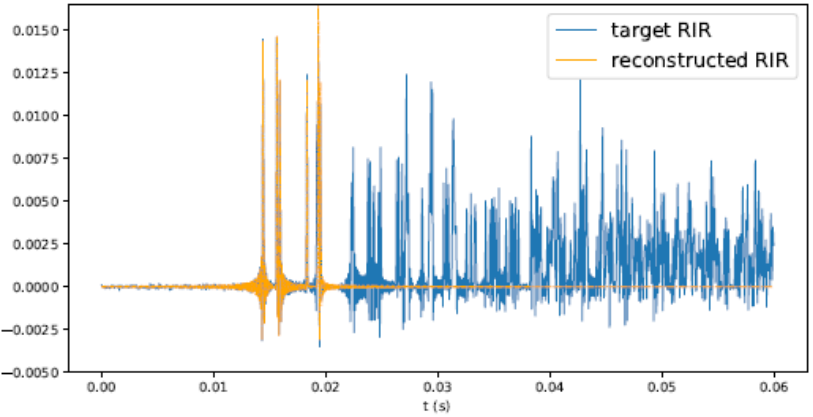


Sliding-Frank-Wolfe [3], iteration 7

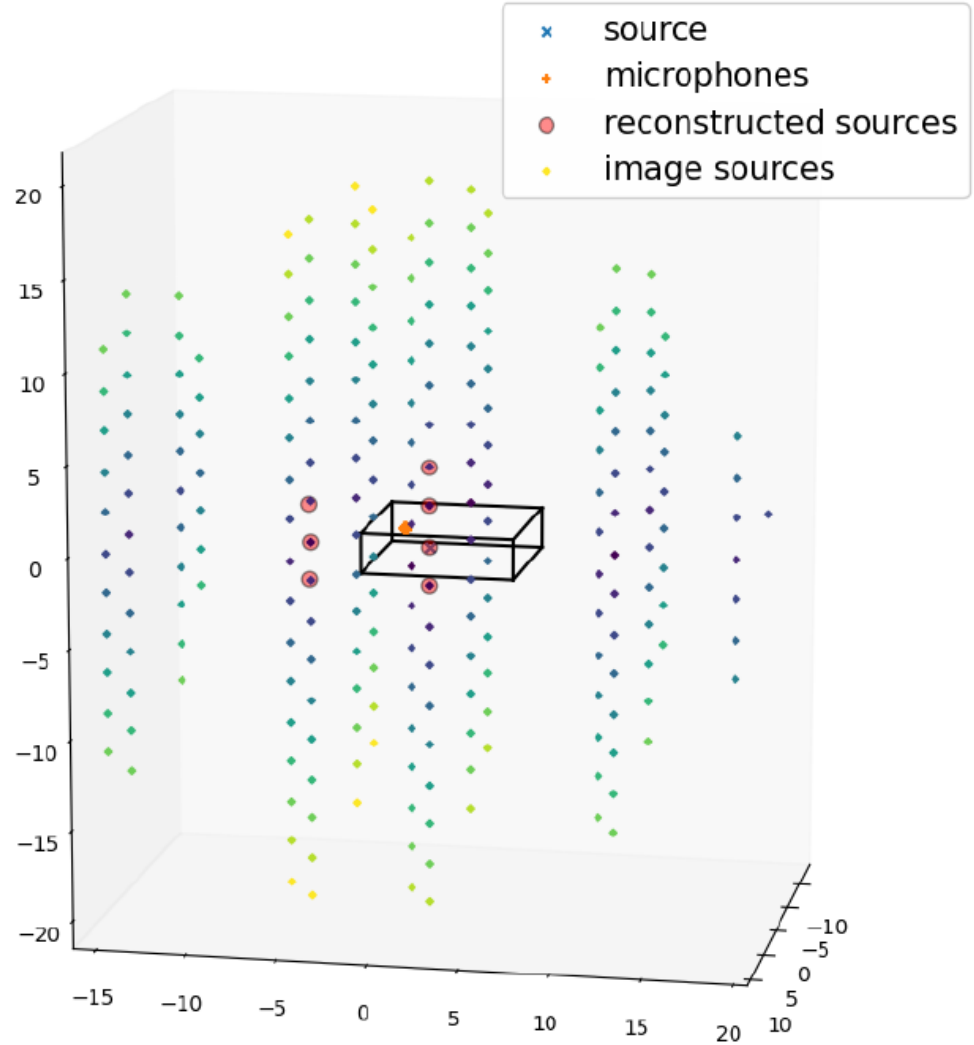
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

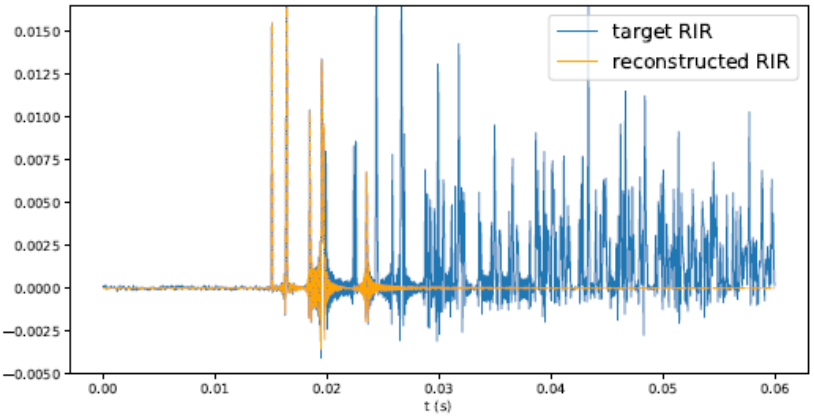


Microphone 2

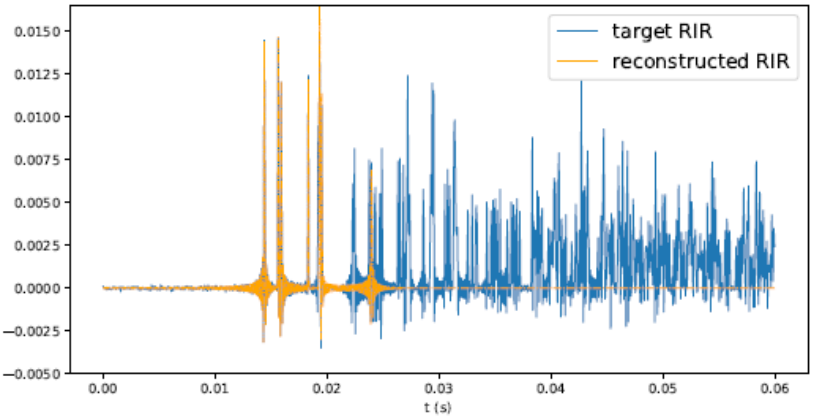


Sliding-Frank-Wolfe [3], iteration 8

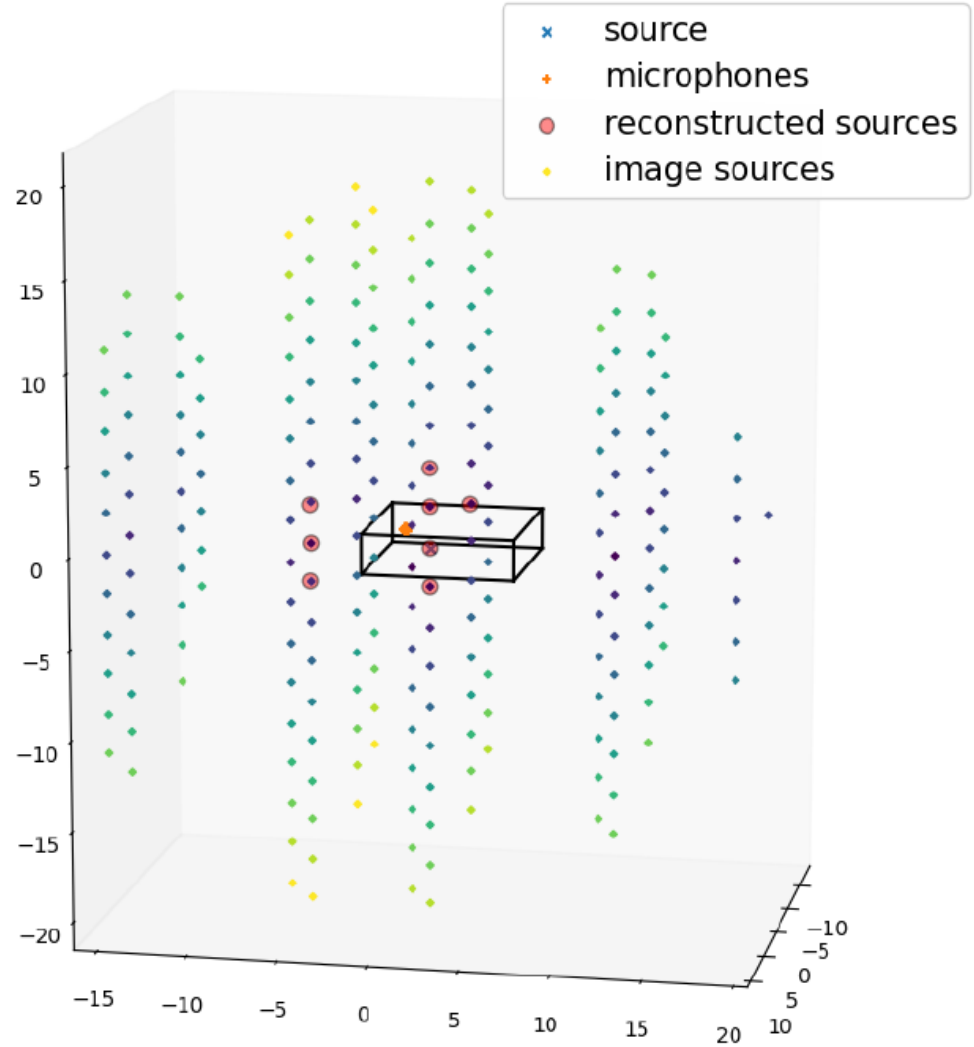
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

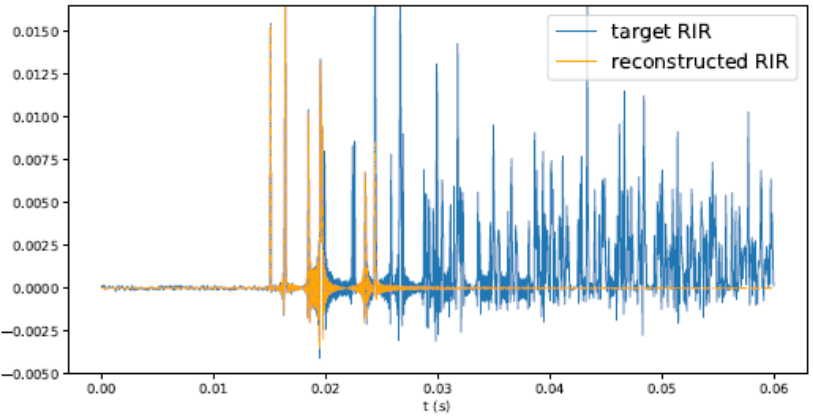


Microphone 2

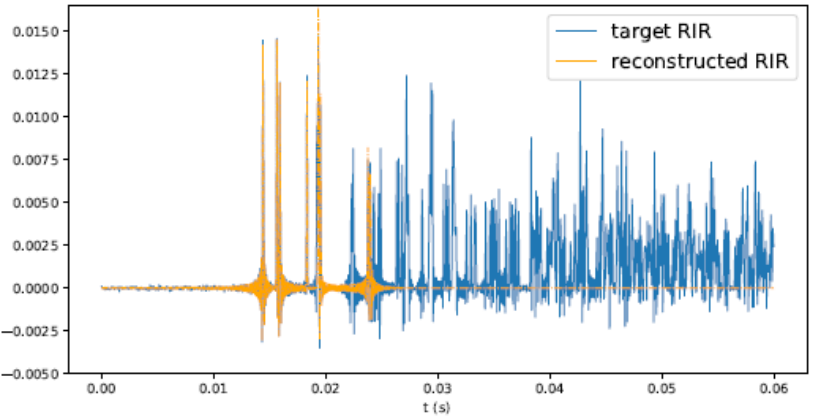


Sliding-Frank-Wolfe [3], iteration 9

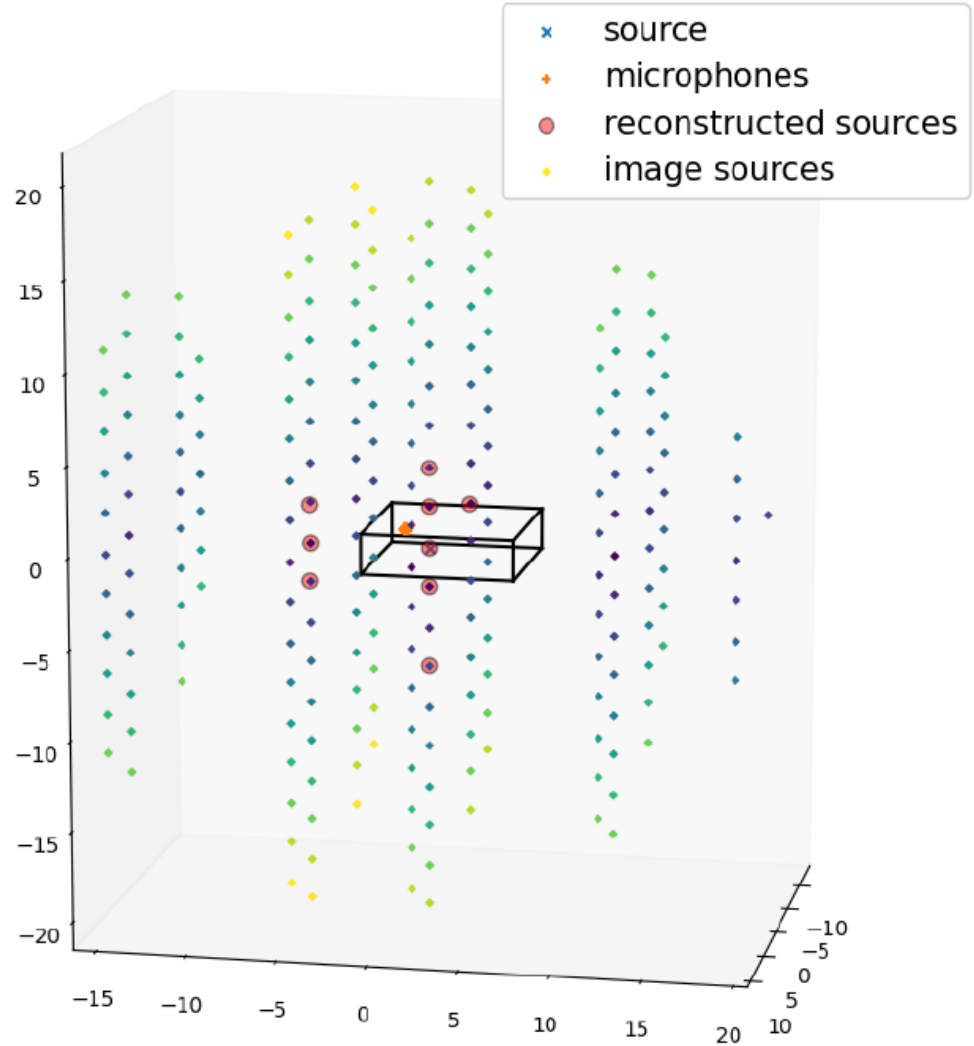
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

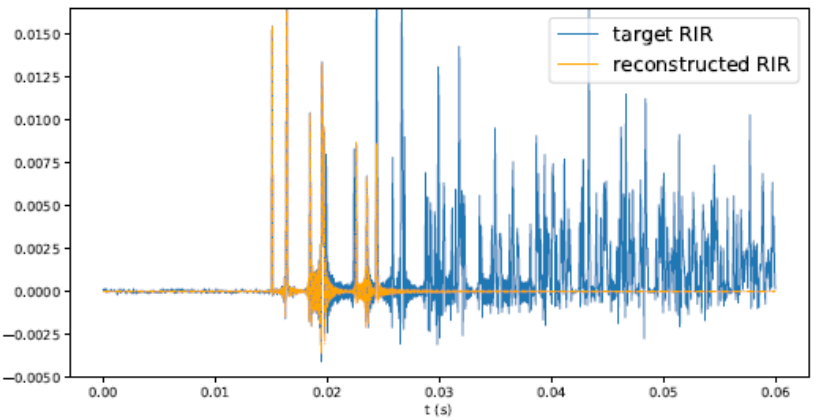


Microphone 2

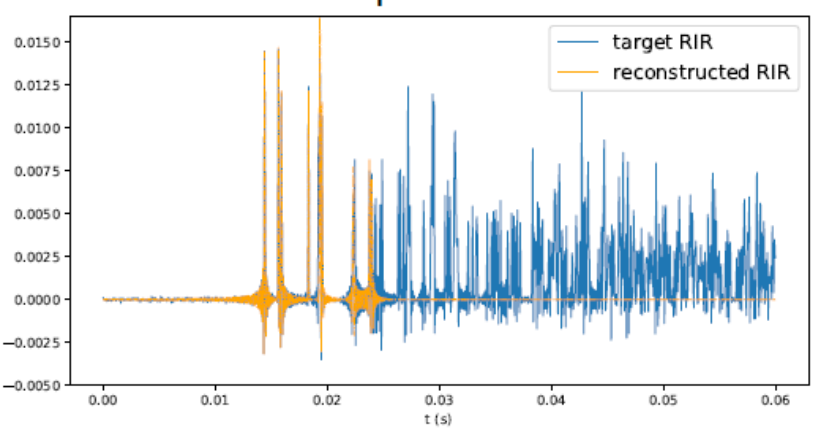


Sliding-Frank-Wolfe [3], iteration 10

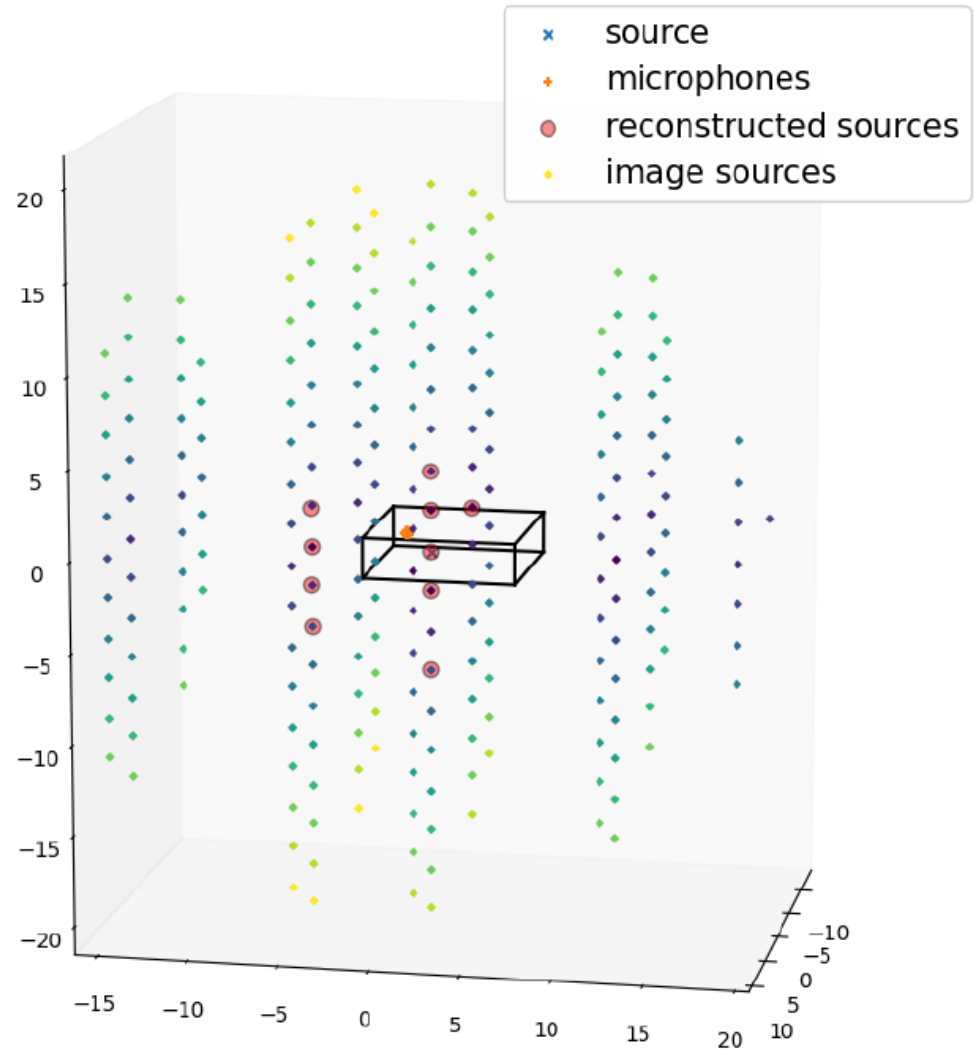
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

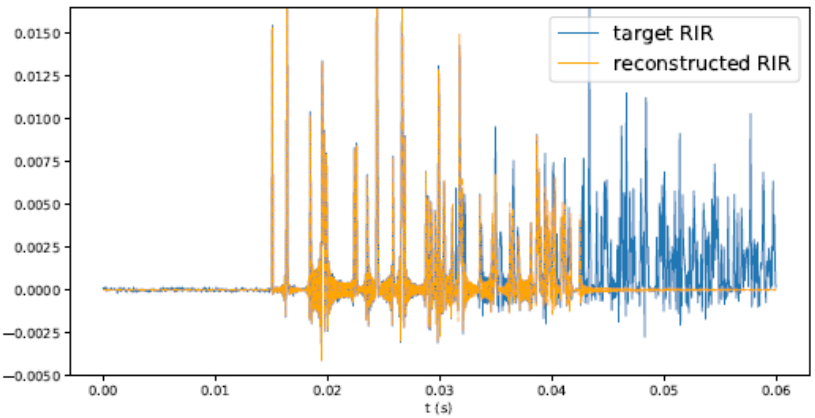


Microphone 2

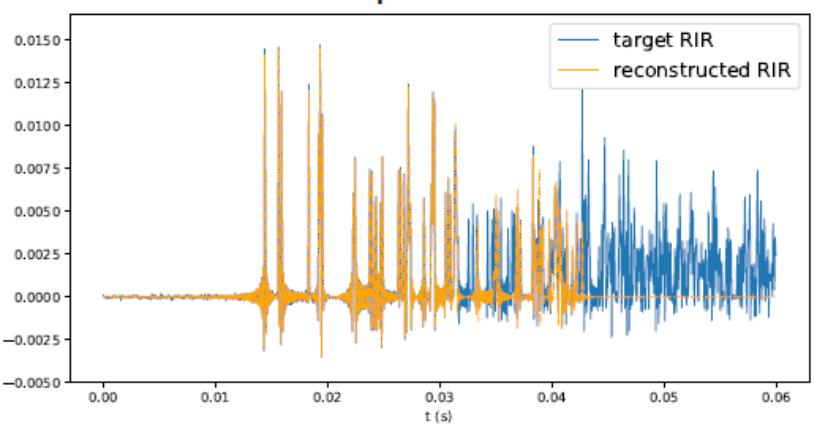


Sliding-Frank-Wolfe [3], iteration 64

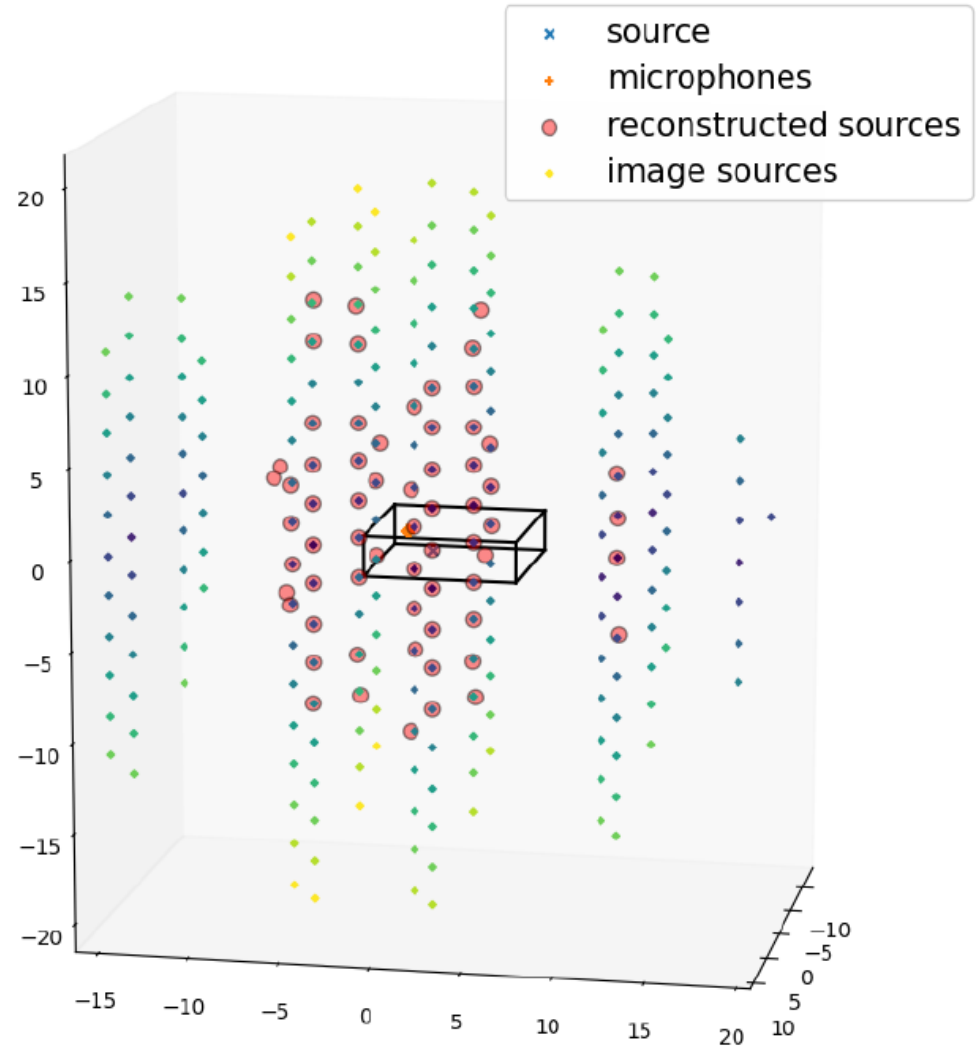
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

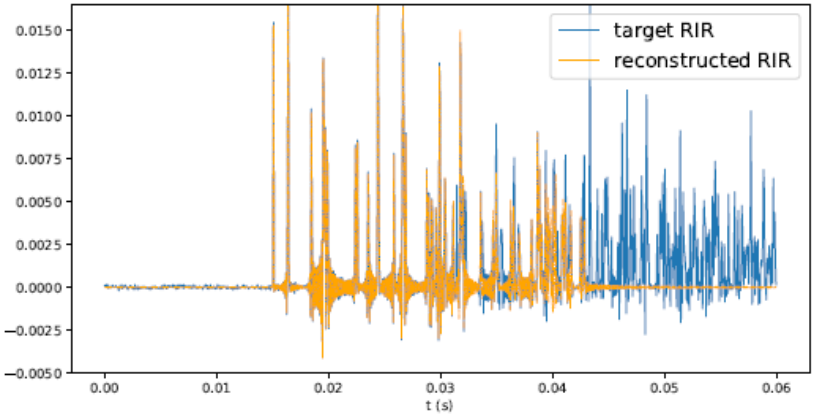


Microphone 2

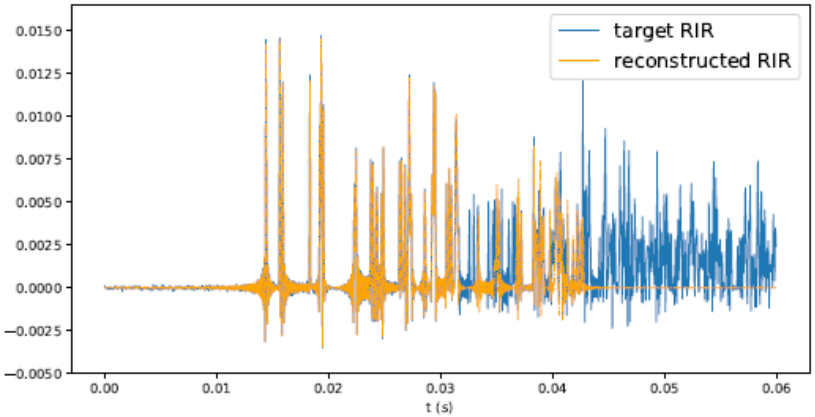


Sliding-Frank-Wolfe [3], iteration 65

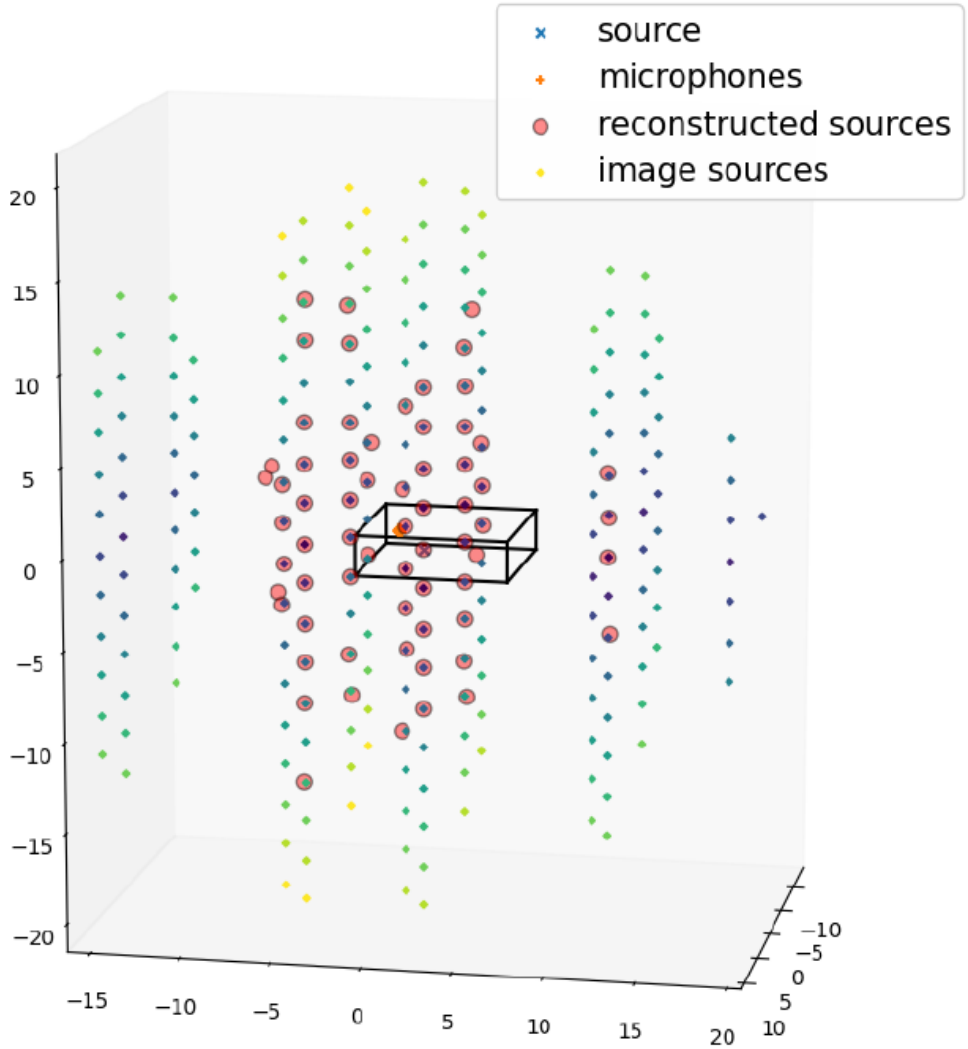
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

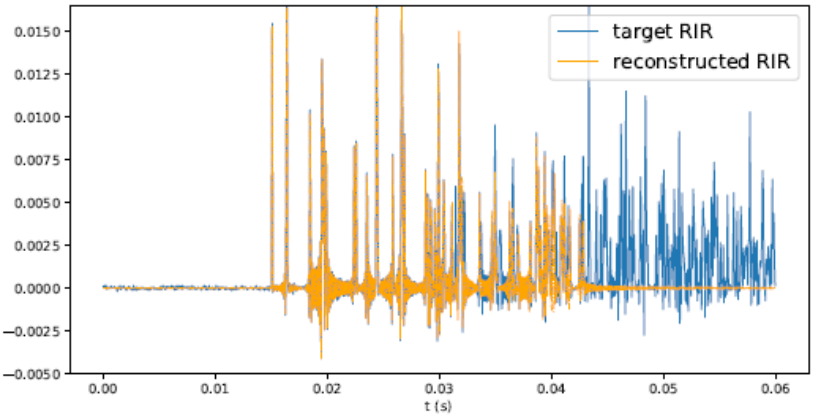


Microphone 2

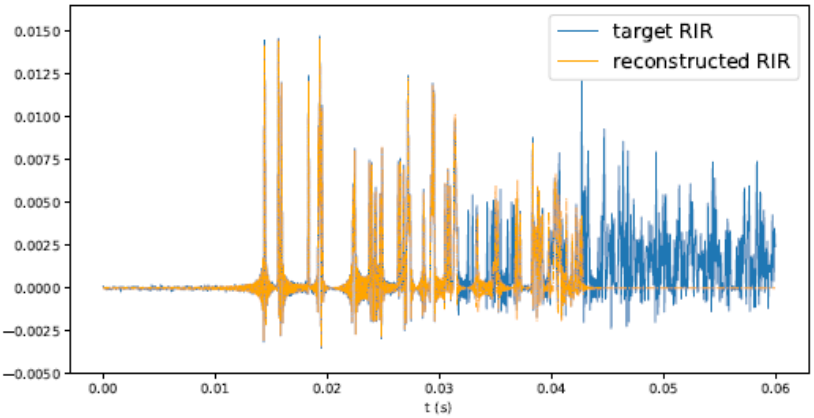


Sliding-Frank-Wolfe [3], iteration 66

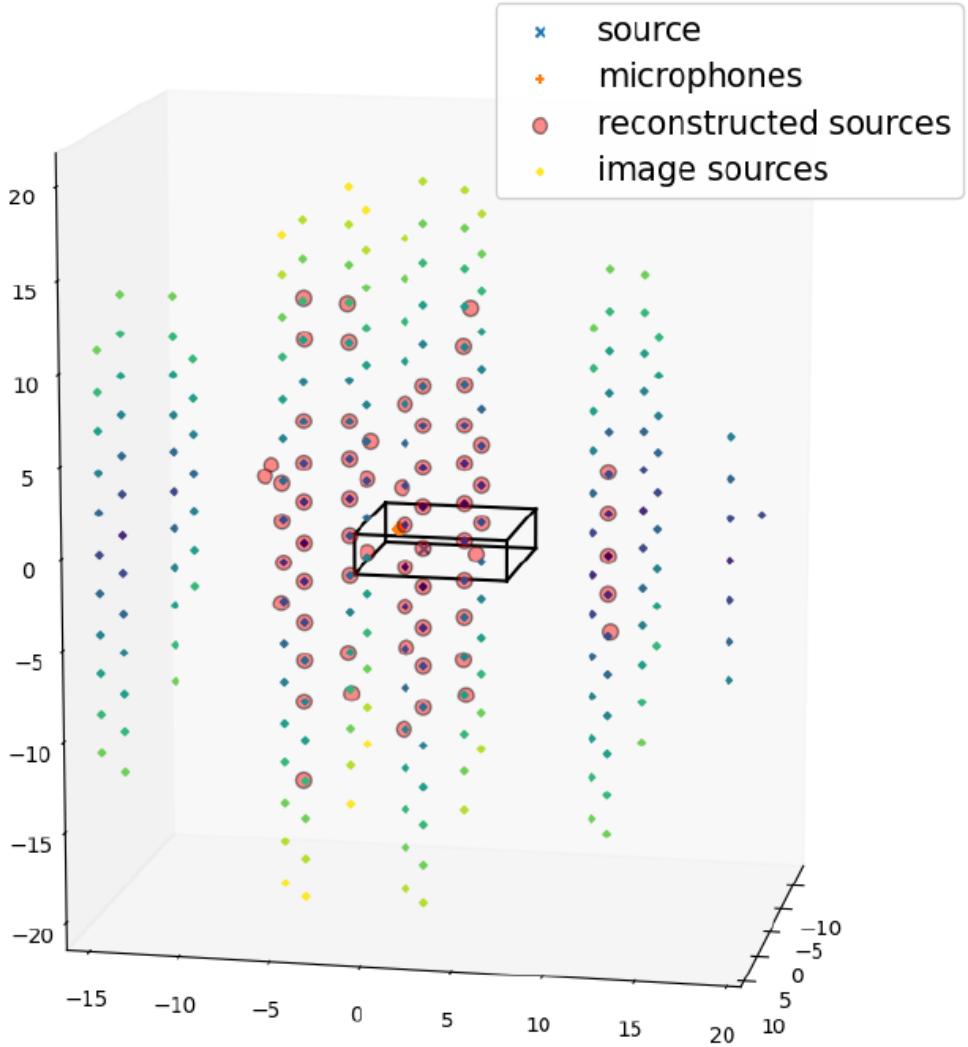
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

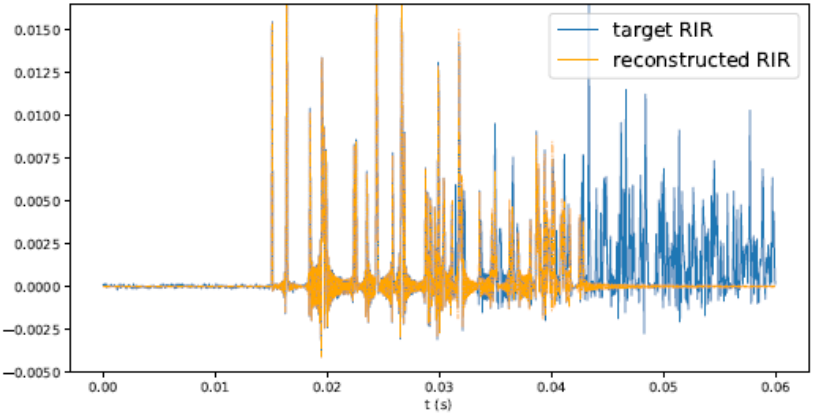


Microphone 2

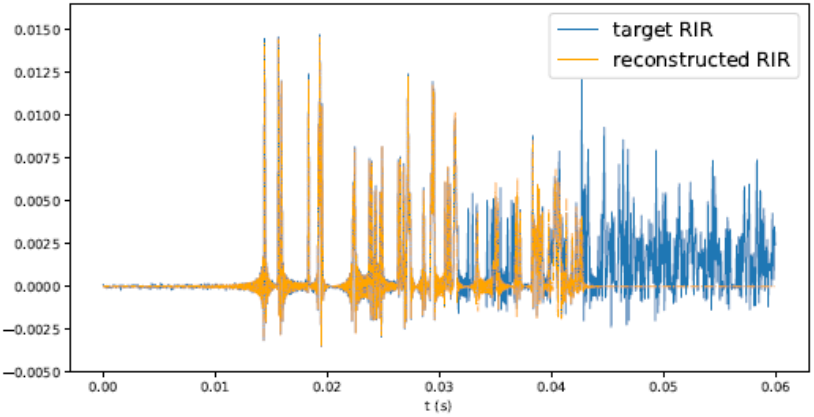


Sliding-Frank-Wolfe [3], iteration 67

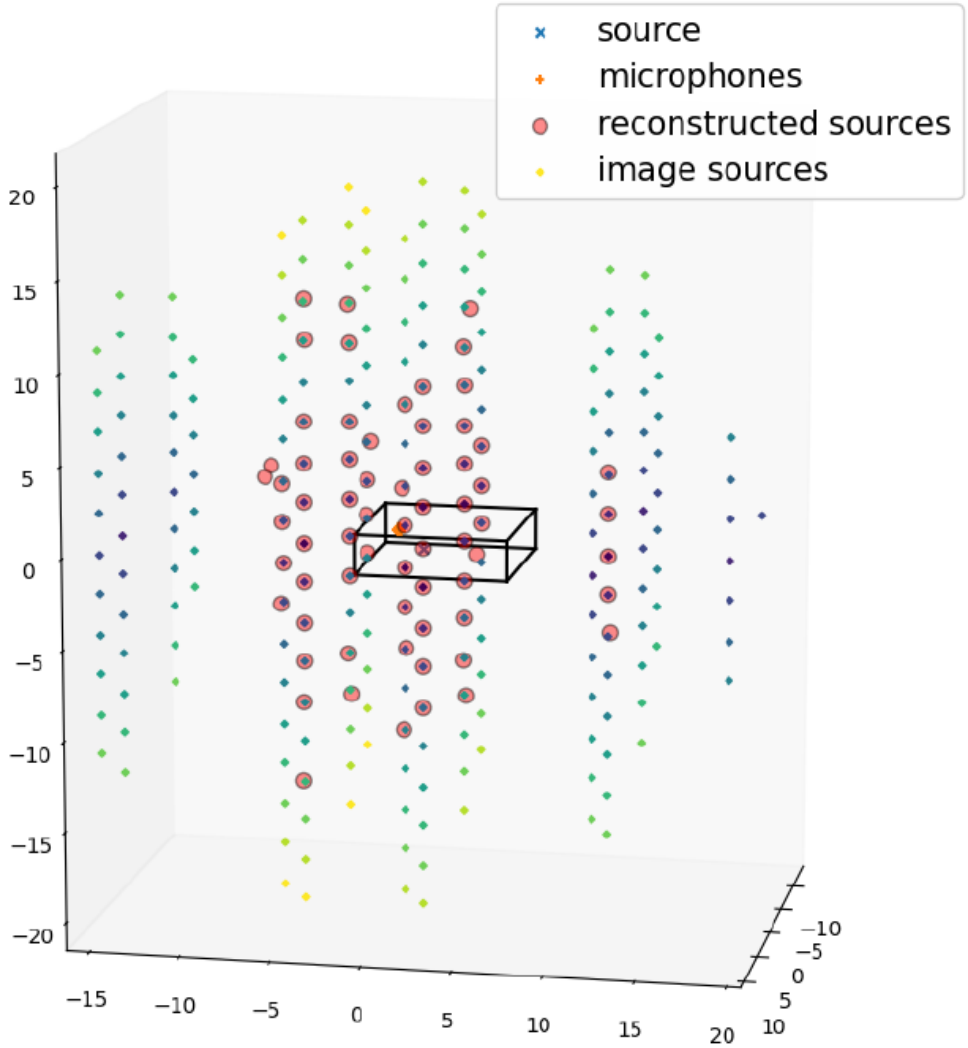
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

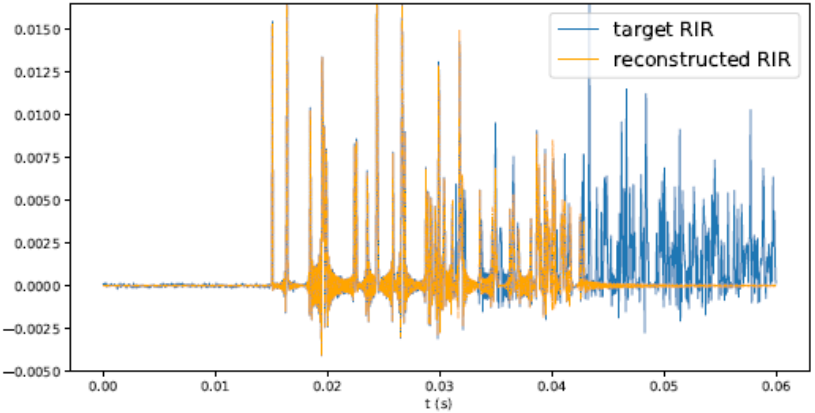


Microphone 2

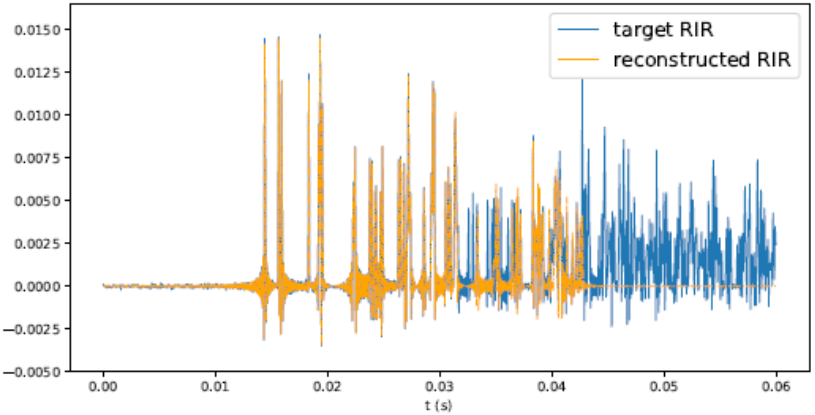


Sliding-Frank-Wolfe [3], iteration 68

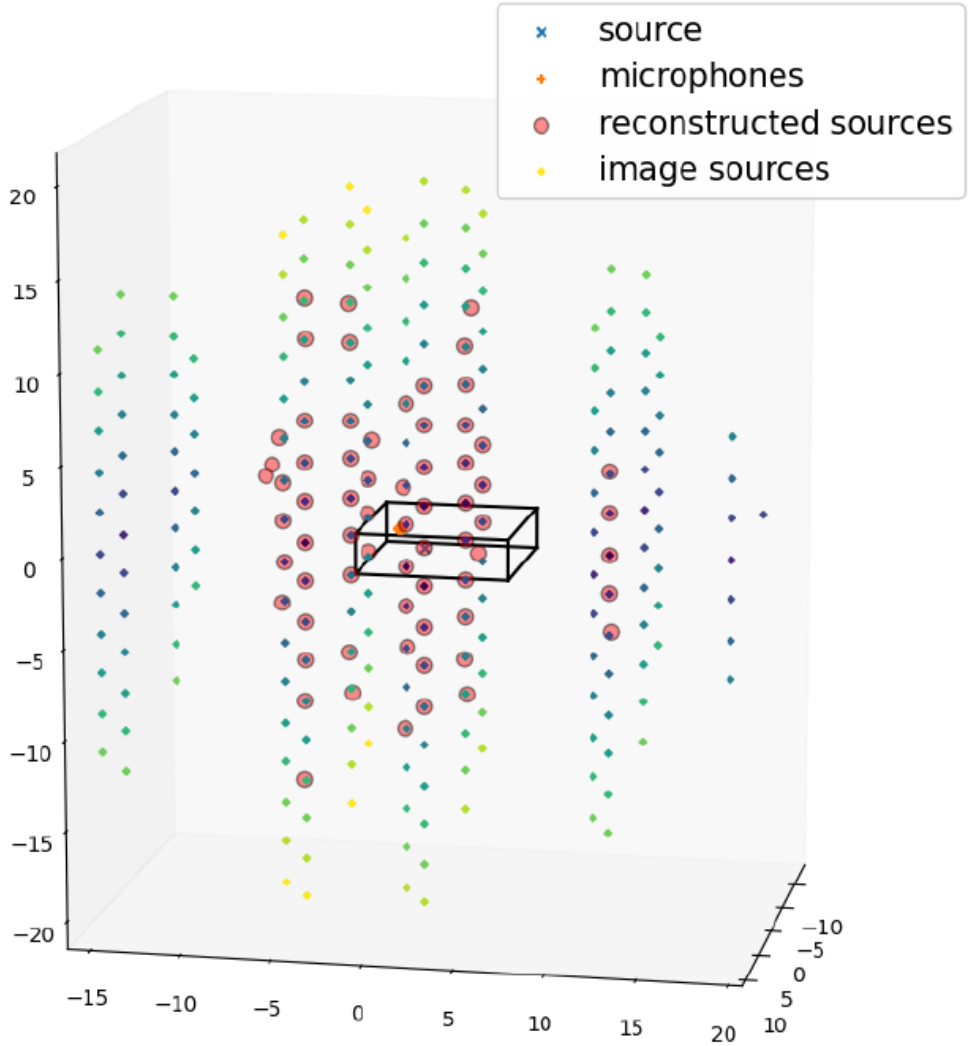
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

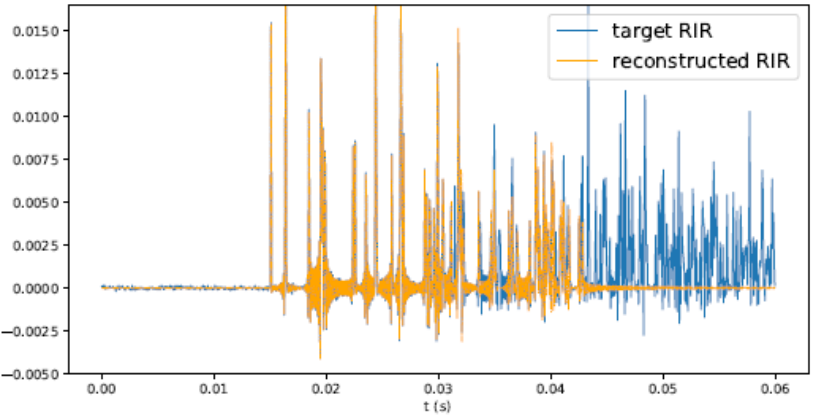


Microphone 2

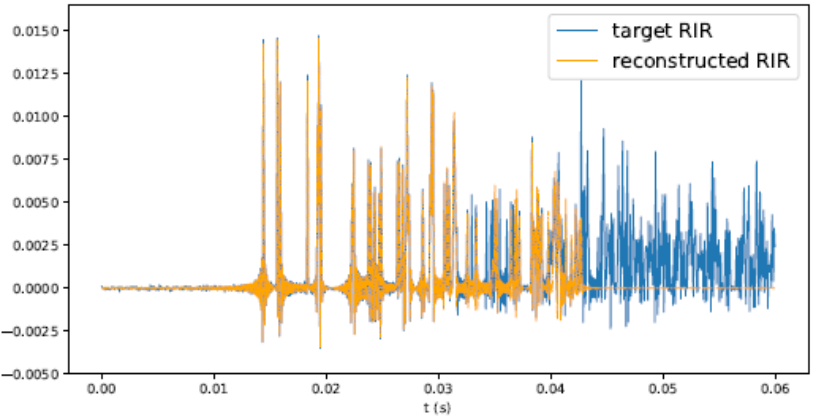


Sliding-Frank-Wolfe [3], iteration 69

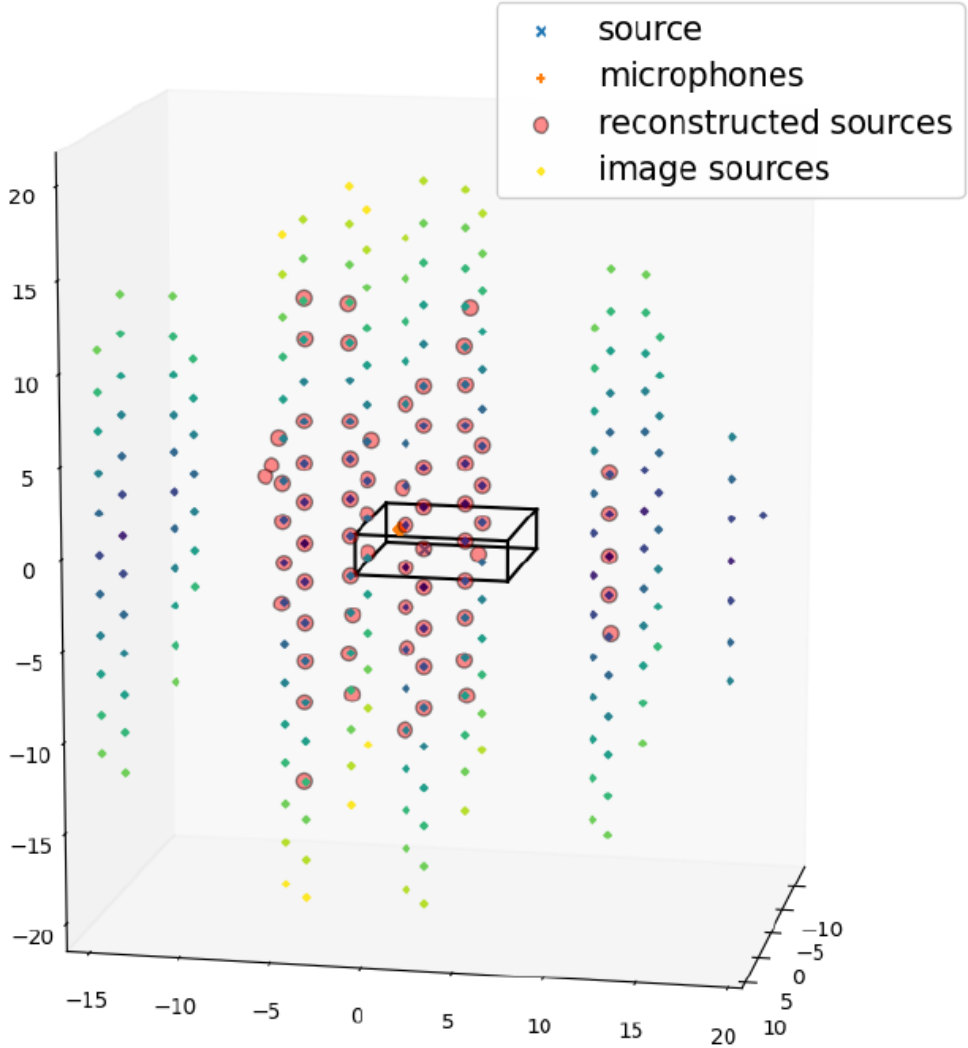
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

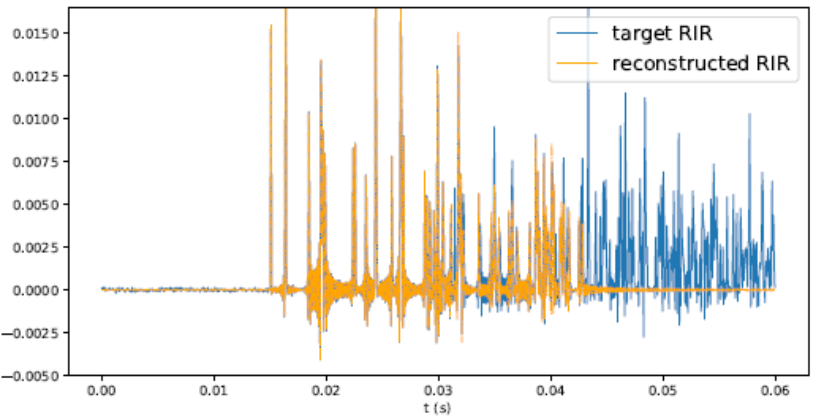


Microphone 2

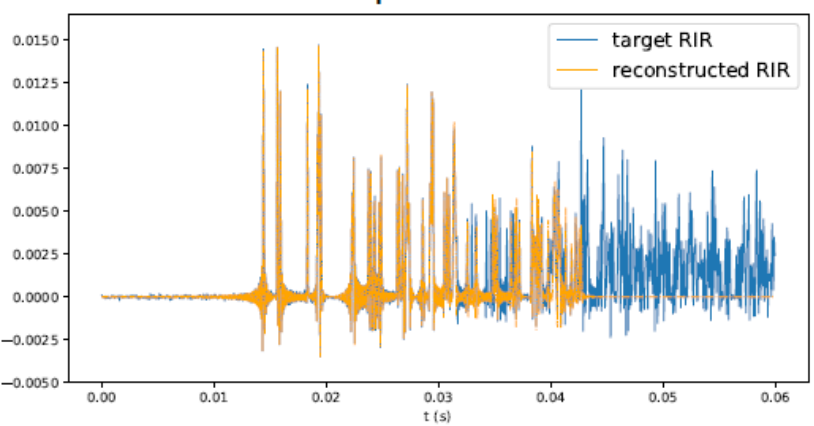


Sliding-Frank-Wolfe [3], iteration 70

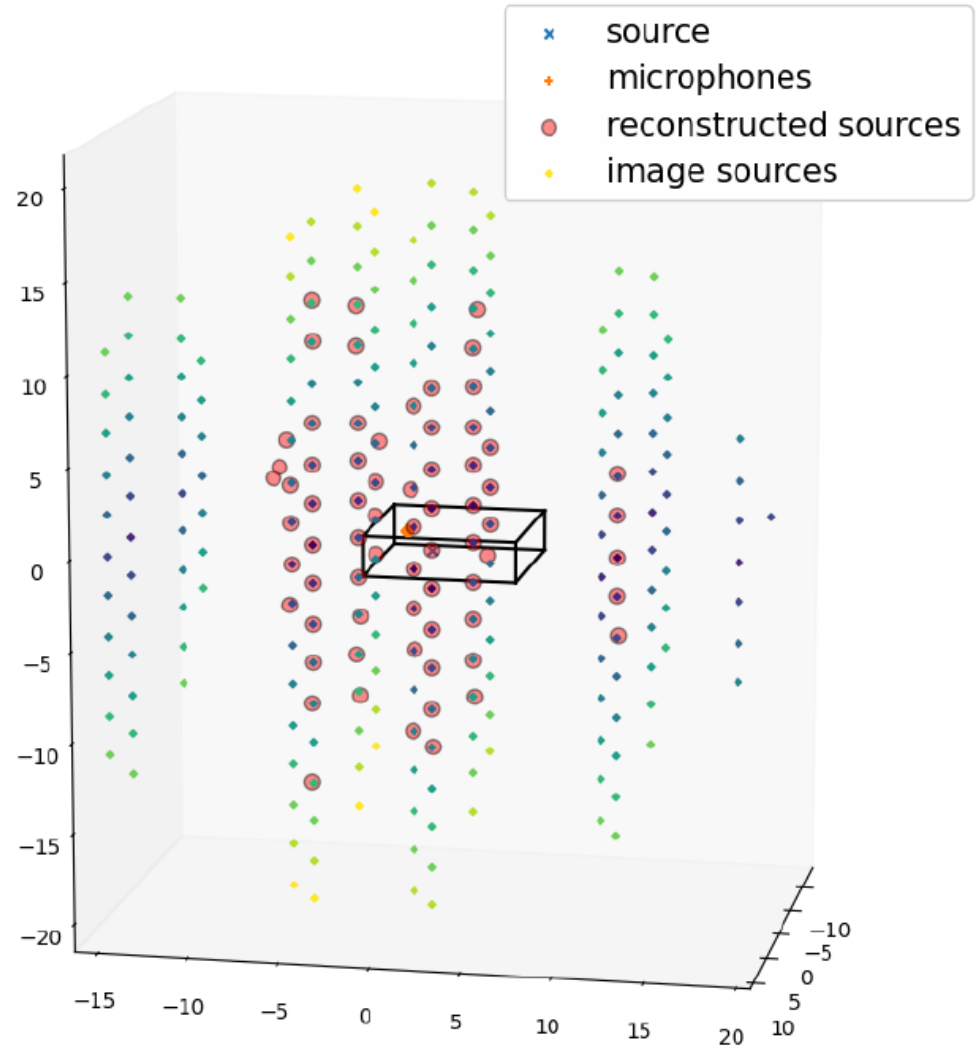
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

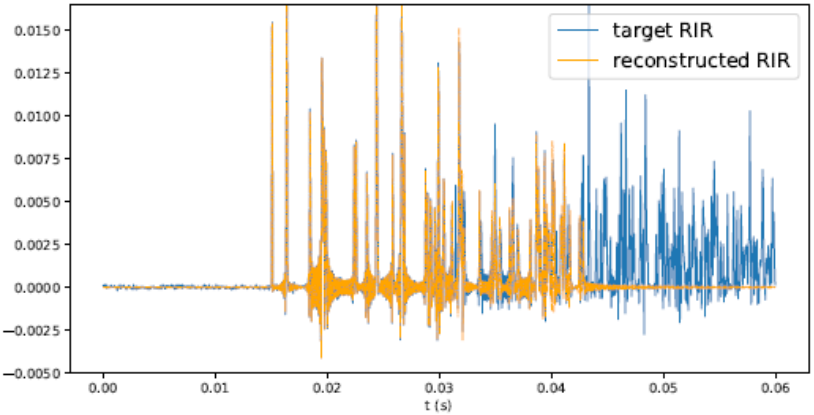


Microphone 2

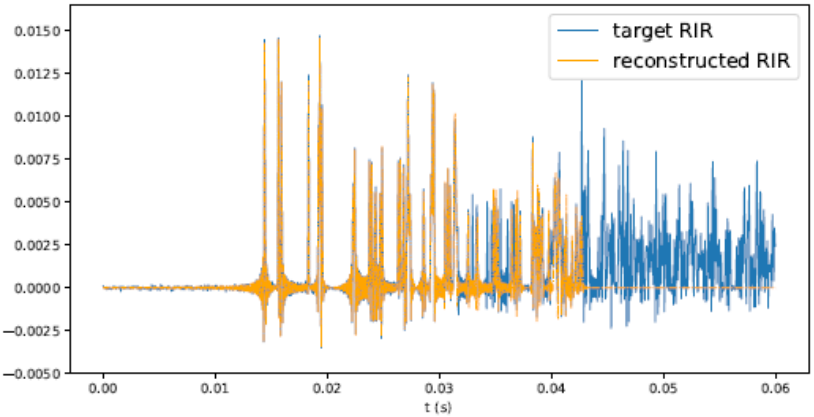


Sliding-Frank-Wolfe [3], iteration 71

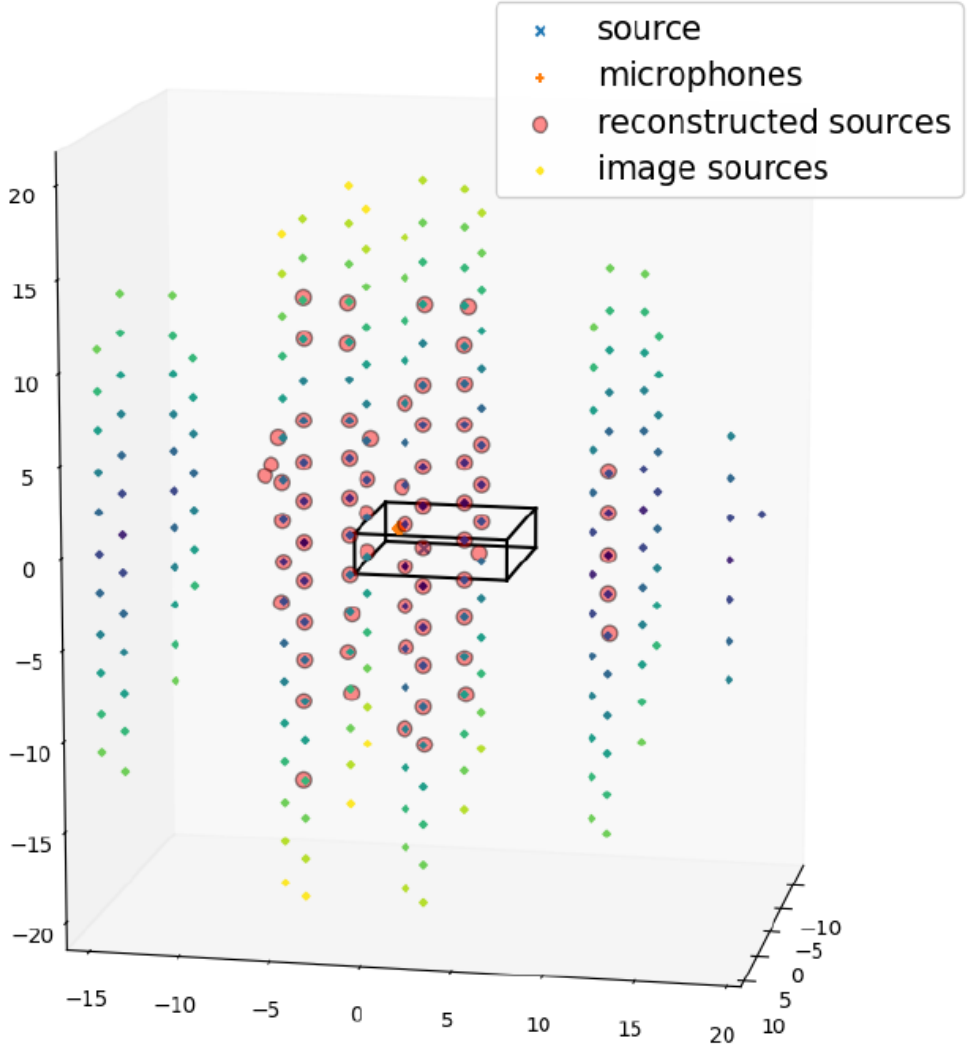
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

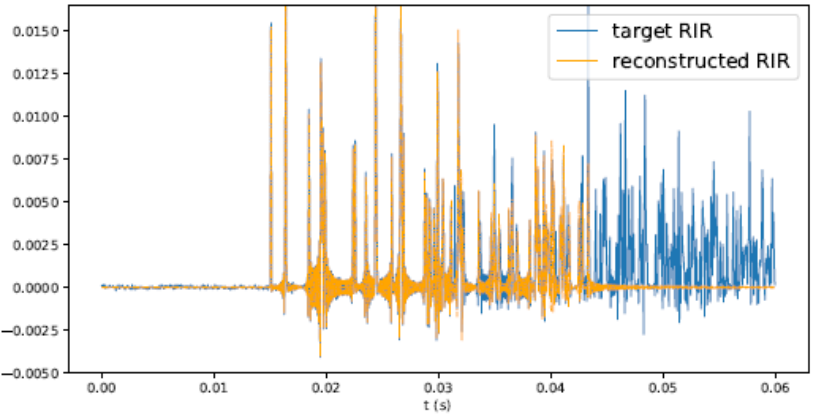


Microphone 2

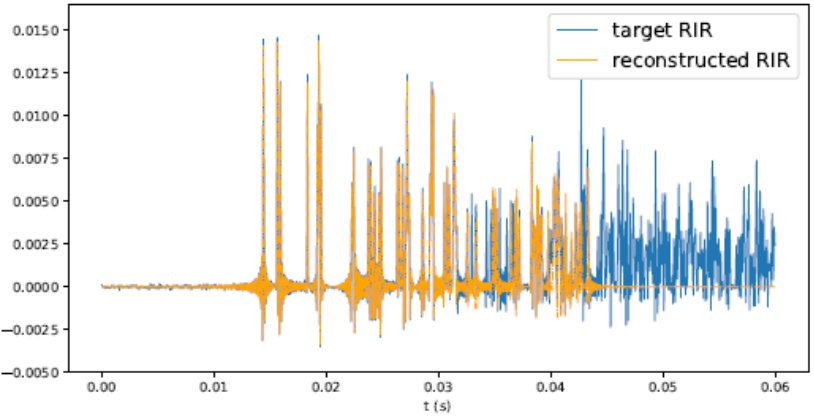


Sliding-Frank-Wolfe [3], iteration 72

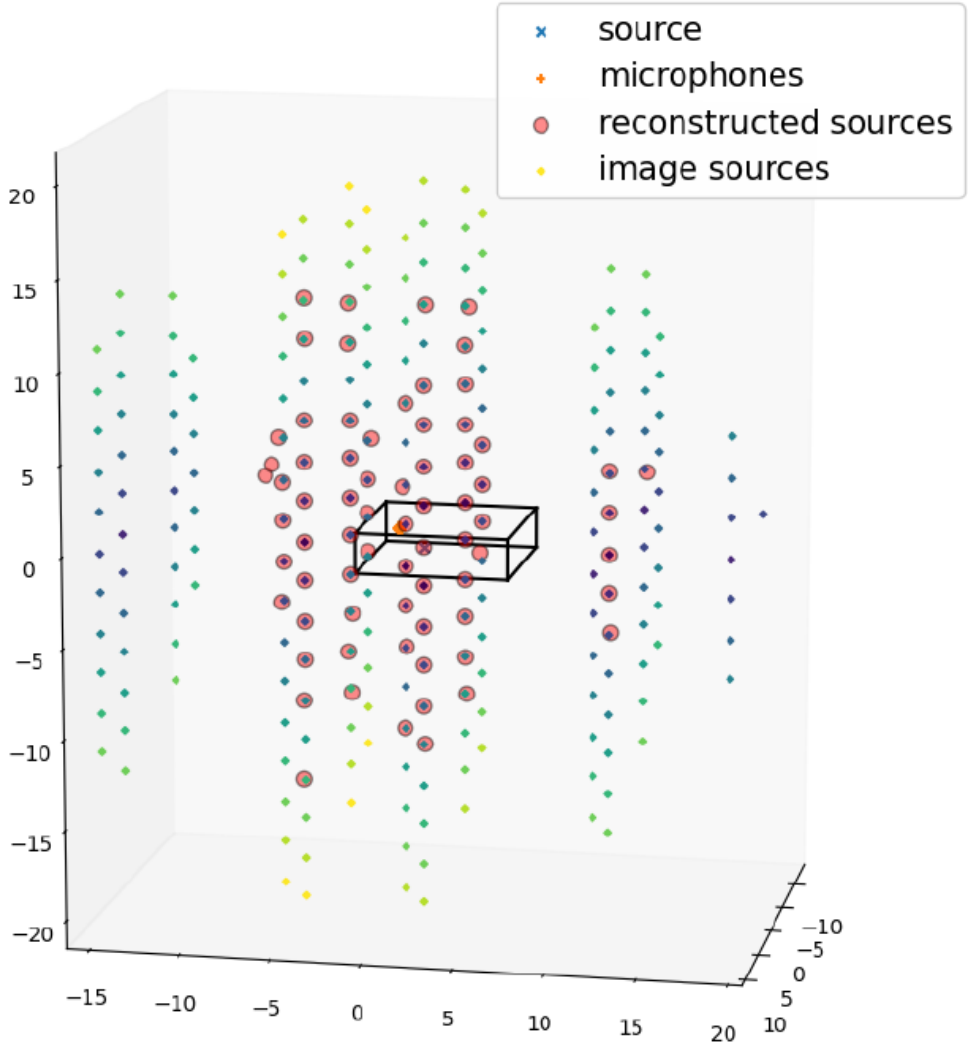
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

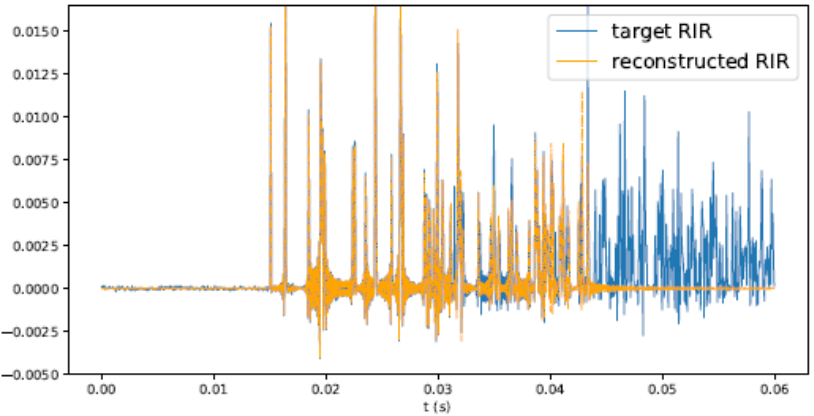


Microphone 2

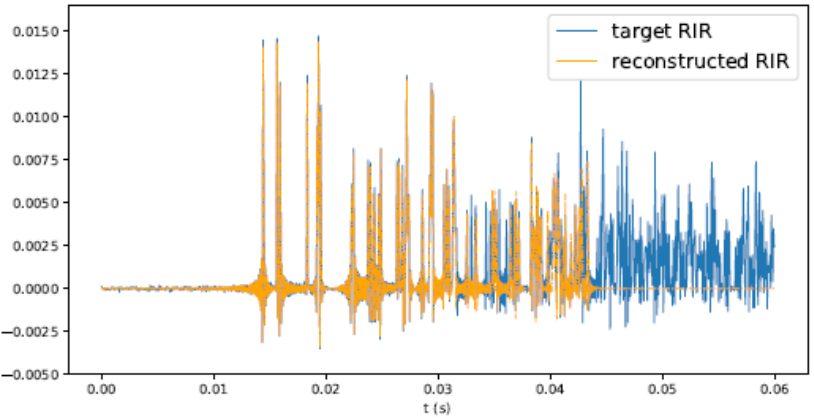


Sliding-Frank-Wolfe [3], iteration 73

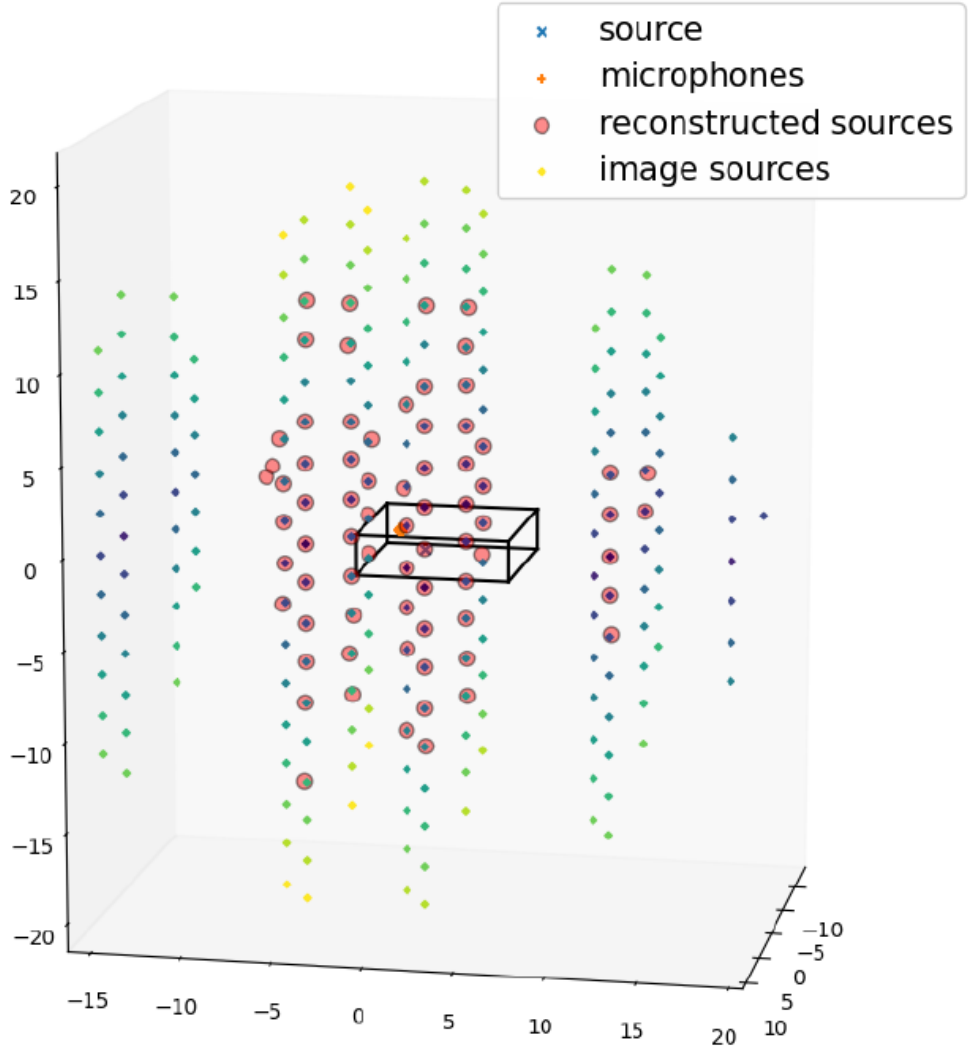
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

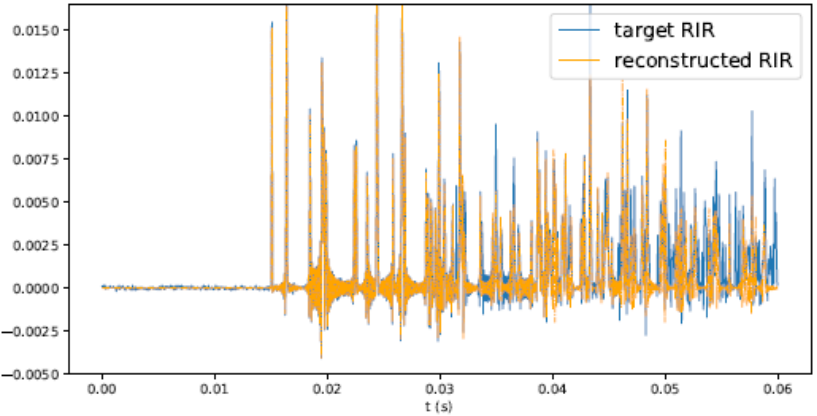


Microphone 2

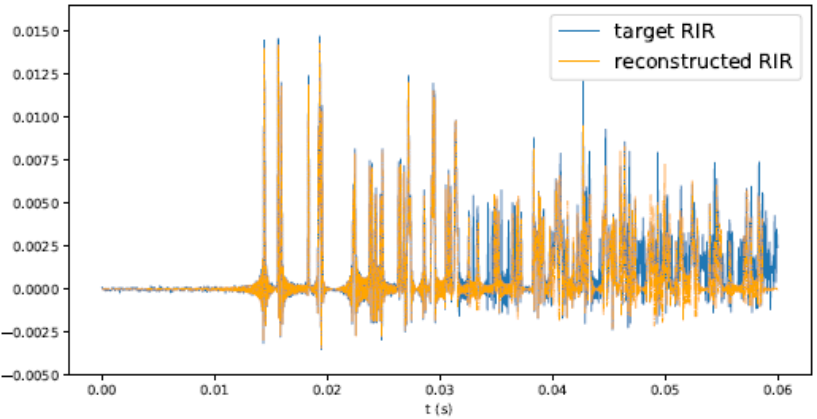


Sliding-Frank-Wolfe [3], iteration 128

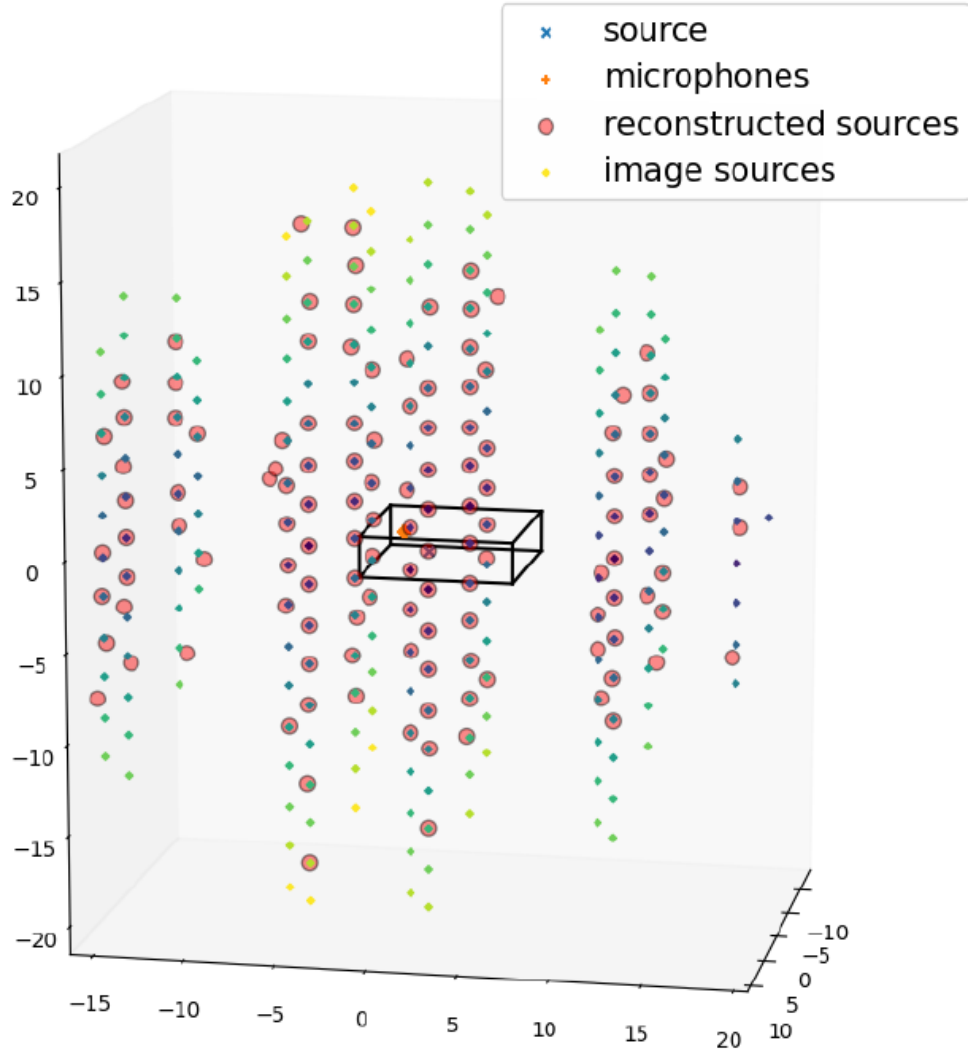
Idea : construct a new Dirac at each iteration based on the residual



Microphone 1

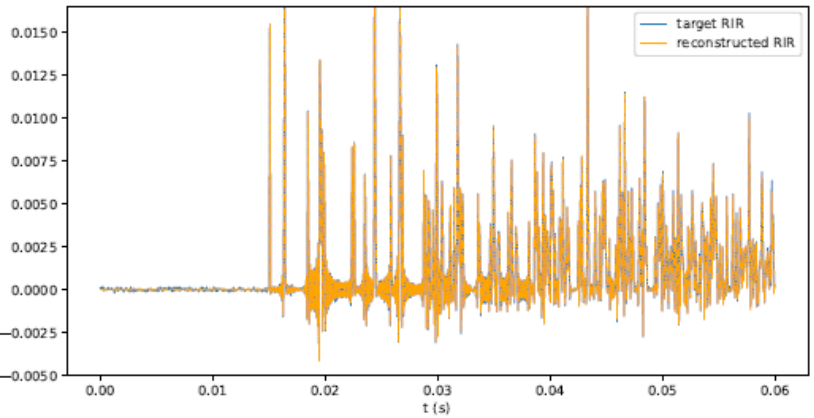


Microphone 2

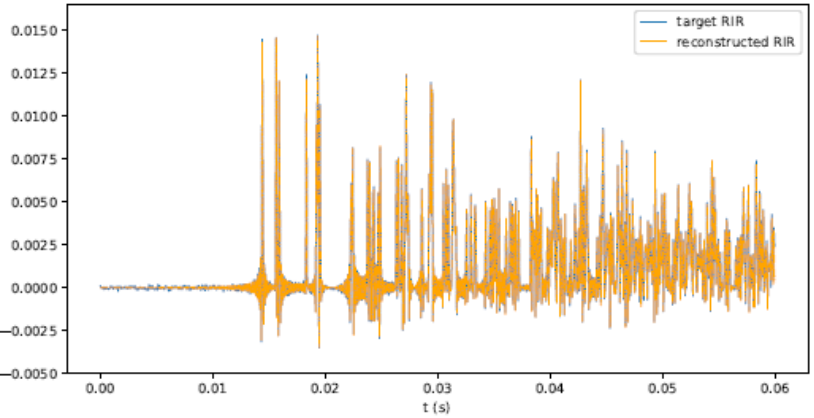


Sliding-Frank-Wolfe [3], iteration 292

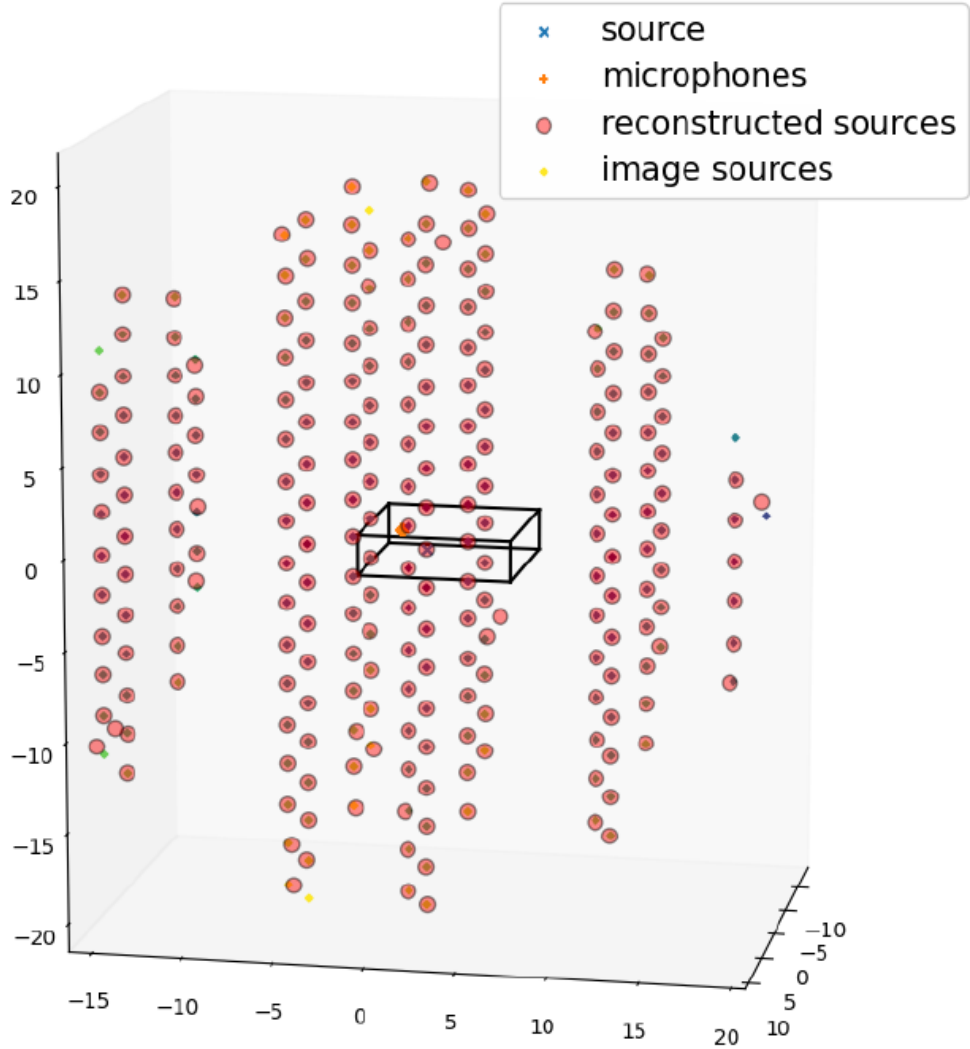
Idea : construct a new Dirac at each iteration based on the residual



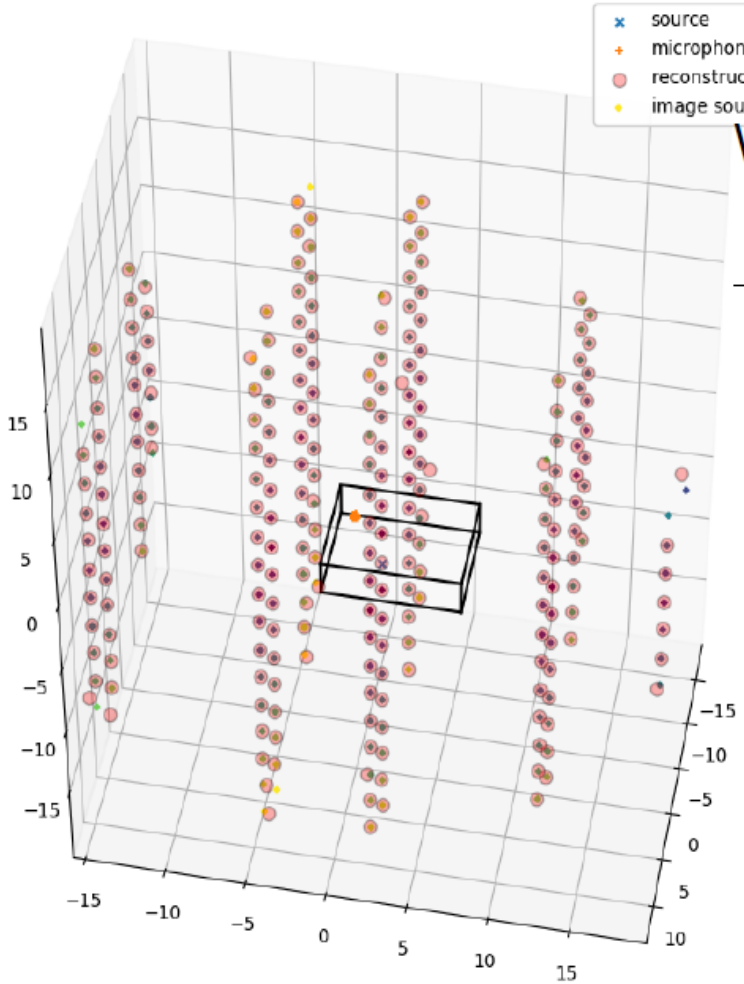
Microphone 1



Microphone 2



Example of reconstruction : 292 iterations for 235 sources ($8.7 \times 8.4 \times 2.2$m room)



With a 5 cm error threshold :

source order	recall	mean error (mm)
1	100%	3
2	100%	6
3	93%	11
4	85%	11

Overall :

- 169/235 (72%) sources recovered at 5cm
- 222/235 (94%) source recovered at 30cm

Ensemble d'observations

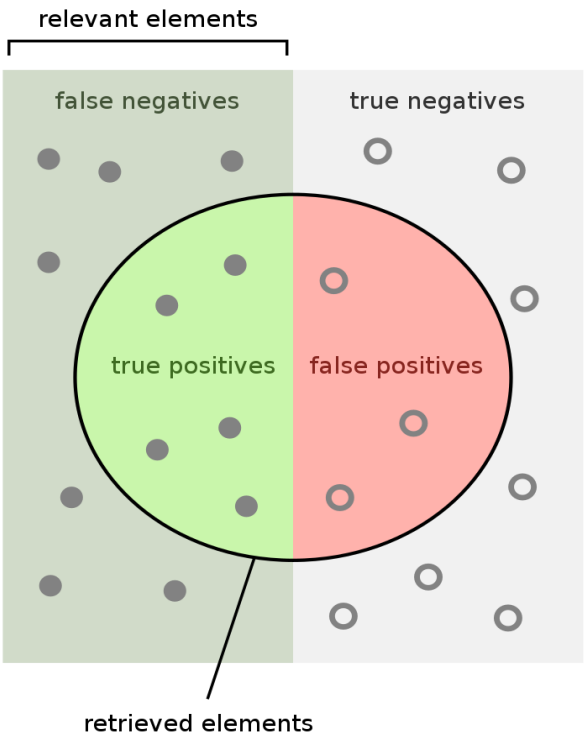
- Données simulées
- 100 salles de dimensions comprises entre : $2\text{m} < L_x < 10\text{m}$, $2\text{m} < L_y < 10\text{m}$, $2\text{m} < L_z < 5m$
- Position de la source et du microphone aléatoires (contrainte : à plus d' 1 m des parois)
- Absorption des parois fixée à 0,1
- Source : antenne sphérique de rayon r (4,2cm, 8,4cm…)
- $\mathbf{Y}(\boldsymbol{X})$: cas idéal, données simulées (python) $\mathbf{Y}(\boldsymbol{X})$ + Bruit gaussien
- Durée des RIR : 60 ms (f_e=16kHz)
- Algo SFW : 128 itérations

Résultats

TP (vrai positif) : source-image construite à moins de 5cm de la source-image théorique correspondante.

FN (faux négatif) : source-image théorique non reconstruite (soit pas de SI, soit SI supérieure à 5cm).

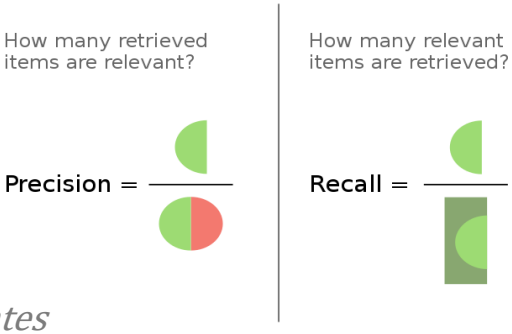
FP (faux positif) : source-image construite sans source-image théorique correspondante à moins de 5cm.



Combien de sources-images trouvées sont pertinentes ?

Proportion des SI bien reconstruites parmi les SI reconstruites

$$Precision = \frac{TP}{TP+FP}$$



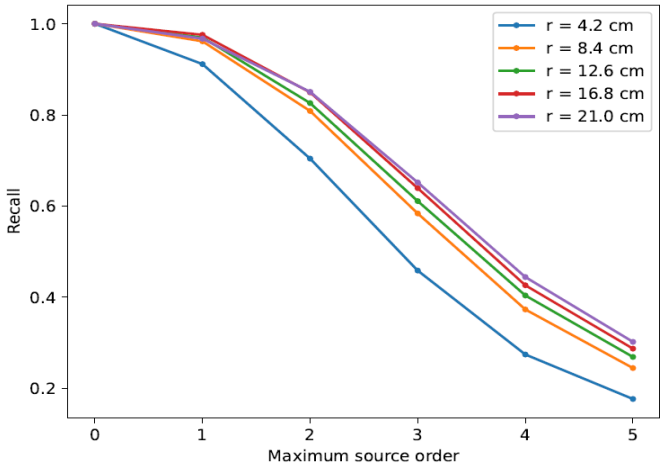
Combien de sources-images pertinentes sont trouvées ?

Proportion des SI bien reconstruites par rapport à toutes les SI théoriques existantes

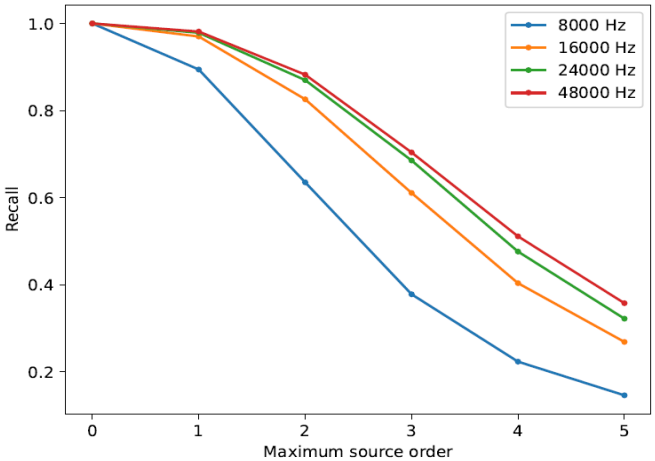
$$Rappel = \frac{TP}{TP+FN}$$

Résultats

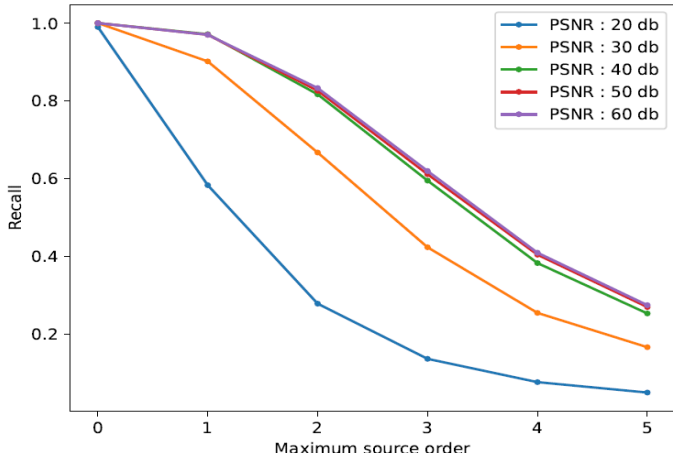
Rayon de l'antenne microphonique



fréquence d'échantillonnage



Bruit



- Evidemment, plus on cherche à reconstruire un grand nombre de SI plus le rappel va chuter.
- Plus la distance entre 2 micros augmente, plus le rappel augmente.
- Plus la fréquence d'échantillonnage augmente, meilleur est le rappel.
- Une bonne robuste au bruit si PSNR supérieur ou égal à 40 dB.

Conclusions & Perspectives

Ca marche pour des RIRs simulées !

Prochaine étape : passage aux RIRs mesurées

Merci