

Description de l'algorithme d'upmix Auralys

Informations générales :

Auteur :

Damien Jacquet – Ingénieur R&D Audio / Fondateur Auralys Audio

Date de rédaction :

13 octobre 2025

Statut du document :

Confidentiel – Diffusion restreinte

Destinataires autorisés :

Équipe R&D Auralys Audio, partenaires techniques et juridiques sous NDA.

Objet :

Ce document décrit les principes, équations et procédés de traitement du signal constituant l'algorithme d'upmix Auralys. Il sert de support technique au dépôt de brevet et à l'évaluation scientifique et industrielle de la technologie.

Clause de confidentialité :

Le présent document contient des informations techniques, scientifiques et conceptuelles confidentielles relatives à l'algorithme d'upmix Auralys. Ces informations sont la propriété exclusive de son auteur et sont protégées par le droit de la propriété intellectuelle. Toute reproduction, diffusion, communication à des tiers, exploitation ou utilisation, totale ou partielle, sans autorisation écrite préalable de l'auteur est strictement interdite. Ce document est remis à titre confidentiel dans le seul but d'évaluation technique ou de dépôt de brevet et ne doit pas être conservé, copié ou transmis à des personnes non autorisées.

Table des matières

Informations générales :	1
Contexte :	2
Objectifs	2
Applications :	3

L'invention :	3
Conventions	4
Prérequis et état de l'art	5
Modèle de localisation d'une source spatialisé par vecteur d'énergie étendus (état de l'art)	5
Analyse de la scène sonore	8
La configuration de mixage :.....	8
Estimation de la direction perçue de la source.....	9
Principe de la transposition du panorama	11
Pré-traitements.....	11
Egalisation	11
Filtre de re-corrélation des basses fréquences.	12
Extraction spatiale	13
TDAP - Triangular Distribution Amplitude Panning (avec correction itérative par vecteur d'énergie)	16
Contexte & objectifs	16
Avantages par rapport aux méthodes existantes	16
Hypothèses & conventions.....	16
1.Pré-traitements	17
2.Génération des gains (profil triangulaire)	17
3.Estimination directionnelle (vecteur d'énergie)	18
4.Boucle d'ajustement (contrôle itératif)	18
5. Solution analytique	19
Re spatialisation.....	20

Contexte :

Objectifs

Cette invention vise à permettre la conversion en temps réel ou non d'un signal stéréo ou multicanal dont on connaît les positions des enceintes (5.1, 7.1...) vers un signal multicanal avec un nombre de canaux différents destiné à un système de restitution dont on connaît la configuration (emplacement des enceintes, type d'enceintes, zone d'écoute...). Cet algorithme doit permettre de restituer fidèlement une scène sonore

encodée dans le signal à transformer dans un lieu d'écoute équipé d'un système de restitution différent et adapté à la salle.

Applications :

Cette invention doit permettre à un exploitant de lieu de diffusion de musique (salle de concert, bar, club, théâtre, auditorium, salon...) de diffuser n'importe quelle œuvre sonore sur son système multicanal de manière fidèle.

L'invention :

L'invention consiste en un algorithme de traitement du signal audio prenant en entrée :

- Un signal audio multicanal
- Des métadonnées :
 - o Configuration du système audio initial sur lequel le mixage a été réalisé :
 - Nombre de haut-parleurs
 - Positions des haut-parleurs
 - Position d'écoute de référence
 - Nature des haut-parleurs : notamment les subwoofers pour lesquels les signaux nommés LFE ont un traitement particulier.
 - o Configuration du système audio de destination :
 - Nombre de haut-parleurs
 - Position et orientation des haut-parleurs
 - Nature des haut-parleurs (large bande ou subwoofers)
 - Caractéristiques des haut-parleurs (ouverture à -3 dB)
 - Positions d'écoutes
 - o Des paramètres de contrôles :
 - Ouverture du panorama souhaité
 - Un paramètre de spread permettant de régler l'étalement et la précision de la scène sonore
 - Un paramètre de réglage de la fréquence basse de traitement, un pré-traitement étant appliqué sous cette fréquence
 - Des paramètres temporels : constantes de temps pour lisser le traitement
 - Des paramètres fréquentiels : taille de fenêtre d'analyse, fenêtre de lissage fréquentiel
 - Un paramètre d'atténuation maximale
 - Un paramètre de pente du filtre d'extraction

Le procédé détermine en sortie un ensemble de signaux audios (signal multicanal, un signal pour chaque haut-parleur de la configuration audio de destination).

Le procédé se déploie en 3 étapes :

- Un étage de pré-traitement
- Un étage d'extraction
- Un étage de re spatialisation.

Conventions

Constantes :

- I : nombre de haut-parleurs dans la configuration électroacoustique initiale. Ne sont comptés que les haut-parleurs large bande, les subwoofers ne sont pas considérés
- J : nombre de haut-parleurs dans la configuration électroacoustique finale. Ne sont comptés que les haut-parleurs large bande, les subwoofers ne sont pas considérés
- K : nombre de portions de panorama considérées
- N_{stft} : nombre de fréquences de l'analyse fréquentielle
- N_{win} : nombre de points de la fenêtre d'analyse fréquentielle
- N_{hop} : le pas de trames (nombre d'échantillons entre 2 fenêtres successives)
- F_s : la fréquence d'échantillonnage
- Ω : angle d'ouverture du panorama initial
- Φ : angle d'ouverture du panorama final
- φ_c : angle du centre du panorama final
- $spread$: facteur d'élargissement (1 par défaut)
- f_{xover1} : fréquence du filtre passe bas de re-corrélation des basses fréquences.
- f_{xover2} : fréquence du filtre passe haut de re-corrélation des basses fréquences.

Indices :

- i : indice relatif au numéro d'un haut-parleur de la configuration initiale considéré, ou de son signal associé,
 $i \in [1, I]$
- k : indice relatif à la portion de panorama considéré, ou de son signal
 $k \in [1, K]$
- j : indice relatif au numéro d'un haut-parleur de la configuration finale considéré, ou de son signal associé
 $j \in [1, J]$
- f : indice relatif à la fréquence considérée lors du traitement dans le domaine fréquentiel
 $f \in [1, N_{stft}]$
- t : indice relatif à l'échantillon temporel d'un signal.
- T : indice temporel dans le domaine fréquentiel (indice de la fenêtre temporelle considérée)

Signaux :

- $e_i(t)$: signal du haut-parleur i de la configuration électro-acoustique initiale dans le domaine temporel.
- $s_i(t)$: signal du haut-parleur i de la configuration électro-acoustique initiale dans le domaine temporel après l'étage de pré traitement.
- $S_i(f, T)$: signal du haut-parleur i de la configuration électro-acoustique initiale dans le domaine fréquentiel considéré à la fréquence f et sur la fenêtre d'analyse T .
- $q_k(t)$: signal temporel associé à la portion de panorama k
- $r_j(t)$: signal du haut-parleur j de la configuration électro-acoustique finale dans le domaine temporel.

Données géométriques

- $PE_c = \begin{bmatrix} x_c \\ y_c \end{bmatrix}$ Position centrale d'écoute
- χ_k : largeur de spatialisation
-

Prérequis et état de l'art

Modèle de localisation d'une source spatialisé par vecteur d'énergie étendus (état de l'art)

Dans ce modèle, la direction perçue d'une source à un position est calculée en fonction de l'énergie du signal de la source reçu de chaque haut parleur à ladite position. La contribution de chaque haut parleur est représentée par un vecteur dont l'origine est la position d'écoute, pointant vers le haut parleur et dont la norme est l'énergie perçue pour ce seul haut parleur. Les vecteurs sont ensuite sommés et le vecteur résultant noté \vec{r}_E pointe vers la direction perçue de la source sonore à la position d'écoute considérée. On a donc :

$$\vec{r}_E = \frac{\sum_{i=1}^I ((\omega_{\tau,i} \cdot \omega_{dir,i} \cdot \omega_{ac,i} \cdot g_i)^2 \cdot \vec{u}_i)}{\sum_{i=1}^I ((\omega_{\tau,i} \cdot \omega_{dir,i} \cdot \omega_{ac,i} \cdot g_i)^2)} \quad (1)$$

Avec :

- \vec{r}_E = vecteur résultant indiquant la direction apparente de la source
- \vec{u}_i = vecteur unitaire indiquant la direction du haut-parleur i au point d'écoute

$$\vec{u}_i = \begin{bmatrix} x_i - x_L \\ y_i - y_L \end{bmatrix}$$

(x_i, y_i) les coordonnées du haut parleur i
 (x_L, y_L) les coordonnées de la position d'écoute considérée

(2)

- g_i = gain de spatialisation du signal sur le haut-parleur i
- $\omega_{ac,i}$ = atténuation due à la propagation acoustique du haut-parleur i au point d'écoute

$$\omega_{ac,i} = s_{ref} * \frac{d_{ref}}{d_i} \quad (3)$$

avec s_{ref} niveau sonore du haut-parleur mesuré dans l'axe à la distance d_{ref} . Les calculs d'intensité étant relatifs, on peut prendre $s_{ref} = d_{ref} = 1$

- $\omega_{dir,i}$ = atténuation due à la directivité du haut-parleur et l'angle φ_i entre l'axe du haut-parleur et l'auditeur. L'atténuation peut être issue d'une mesure ou approximée à l'aide d'une fonction de directivité générique. Pour simuler la directivité d'un haut-parleur avec un angle d'ouverture dir , on peut prendre par exemple une fonction cardioïde dont l'atténuation à un angle $\frac{dir}{2}$ est équivalente à -6dB :

$$e_{dir} = -\frac{6}{20 \cdot \log_{10} \left(\frac{1 + \cos(dir)}{2} \right)} \quad (4)$$

$$\omega_{dir,i}(\varphi_i) = \left(\frac{1 + \cos \varphi_i}{2} \right)^{e_{dir}} \quad (5)$$

- $\omega_{\tau,i}$ le facteur de pondération en fonction du retard du signal provenant du haut-parleur i à la position d'écoute par rapport au premier évènement auditif.

$$\omega_{\tau,i} = 10^{\tau \cdot \frac{\Delta t_i}{20}} \quad (6)$$

$$\text{avec } \Delta t_i = (\delta_{ac,i} + \delta_{spat,i}) - \min_i (\delta_{ac,i} + \delta_{spat,i}) \quad (7)$$

avec $\delta_{ac,i}$ = retard de propagation acoustique du haut-parleur i à la position d'écoute

$$\delta_{ac,i} = \frac{d_i}{c} \quad (8)$$

- o d_i étant la distance du haut-parleur i au point d'écoute

- $\delta_{spat,i}$ étant le retard de spatialisation sur le haut-parleur i
- $c = 340 \text{ m.s}^{-1}$ la vitesse du son dans l'air
- τ = pente du seuil d'écho. τ Dépend de la nature du signal. Une valeur moyenne de -0.25 dB/ms peut être utilisée

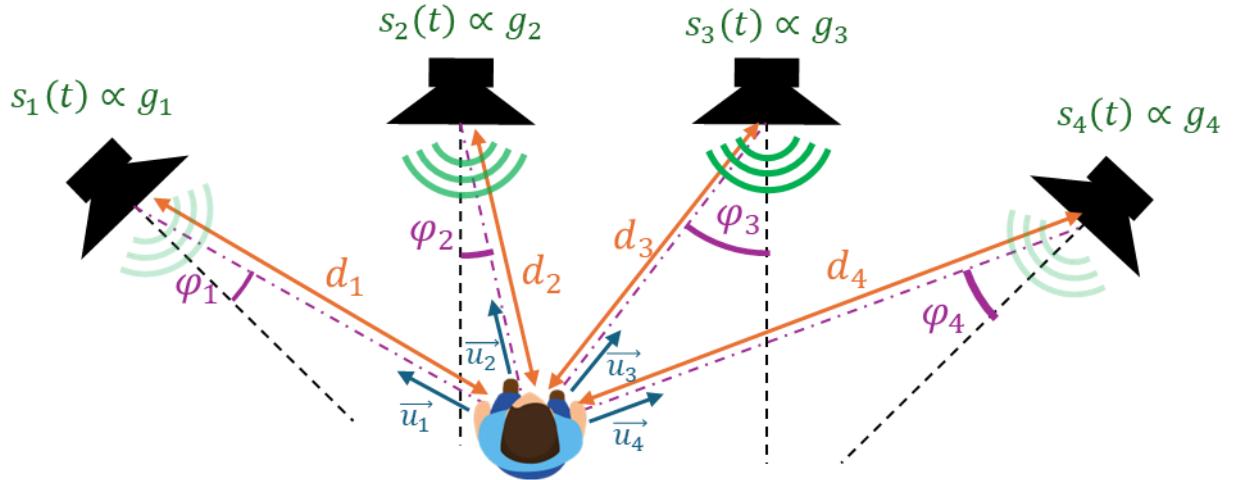


Figure 1: données d'entrée du modèle par vecteur d'énergie

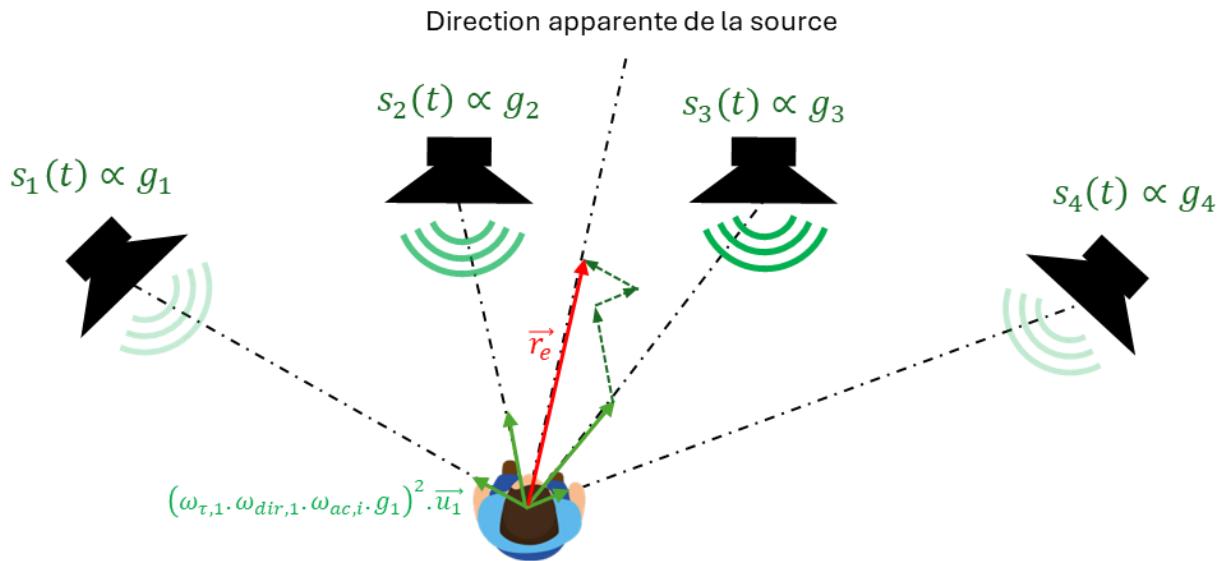


Figure 2 : contribution de chaque haut-parleur et vecteur d'énergie résultant de la sommation.

La largeur perçue de la source peut être calculée comme suit :

$$W = \frac{5}{8} \cdot \frac{180^\circ}{\pi} \cdot 2 \cdot \cos^{-1} \|\vec{r}_E\| \quad (9)$$

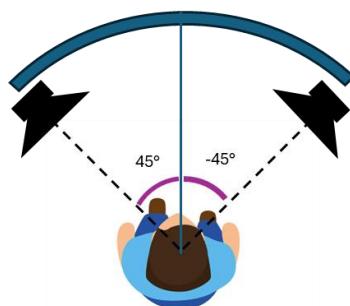
Analyse de la scène sonore

Nous considérons ici uniquement des configurations 2D pour la simplicité des illustrations, la méthode est extensible de manière immédiate à un setup 3D

La configuration de mixage :

L'ingénieur du son chargé de mixer un morceau de musique va choisir une configuration de haut-parleurs : stéréo ou « surround »

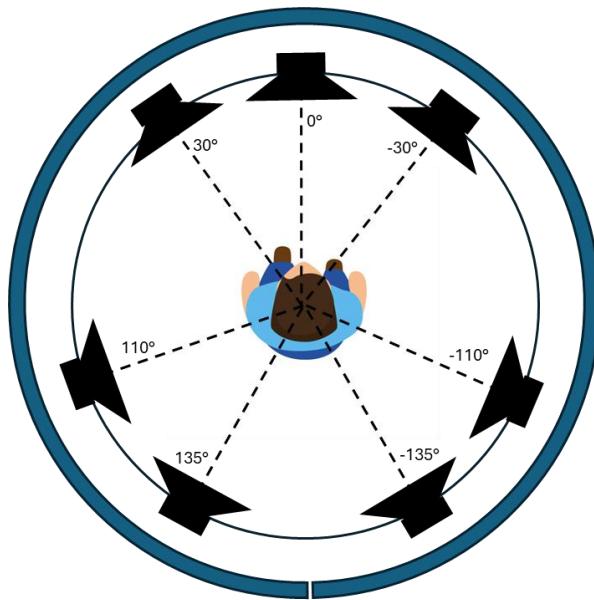
Les configurations stéréos permettent de créer une scène sonore étalée entre les 2 enceintes du système appelée panorama, soit sur une zone couvrante l'angle entre les 2 enceintes



Angle couvert pour le setup stéréo = 90°

Figure 3: illustration d'un setup stéréo

Les configurations « surround » permettent généralement de placer des sources sonores tout autour de l'auditeur, soit un panorama à 360°



Angle couvert pour le setup 7 canaux = 360°

Figure 4: illustration d'un setup multicanal

L'angle d'ouverture maximum de la scène sonore est noté Ω

L'ingénieur du son ayant créé le signal multicanal à transformer a mixé différents signaux sonores issus de la session d'enregistrement de manière à les répartir dans l'espace sur l'ensemble du panorama. Pour répartir les sources sonores, l'ingénieur du son utilise des méthodes de panning qui consistent à diffuser le même signal avec un niveau défini pour chacun des haut-parleurs de manière à donner l'illusion de la localisation du son à un emplacement sur cette zone. De nombreuses méthodes de spatialisation existent, la méthode présentée ici n'en tient pas compte et fonctionne donc avec toute méthode. La position perçue de ce son peut être calculée grâce au modèle par vecteur d'énergie étendu [1].

Estimation de la direction perçue de la source

Le niveau du signal de la source $s_i(t)$ en entrée de chaque haut parleur dépend du gain de mixage de la source g_s , identique pour tous les canaux et du gain de spatialisation g_i propre à chaque haut parleur.

$$s_i(t) = s(t) \cdot g_s \cdot g_i \quad (10)$$

Comme on ne connaît pas le gain de spatialisation utilisé, on peut utiliser directement l'amplitude du signal $s_i(t)$ comme le prouve l'équation 11 pour retrouver le modèle par vecteur d'énergie:

$$\frac{\sum_{i=1}^I \left((\omega_{\tau,i} \cdot \omega_{dir,i} \cdot \omega_{ac,i} \cdot s_i(t))^2 \cdot \vec{u}_i \right)}{\sum_{i=1}^I \left((\omega_{\tau,i} \cdot \omega_{dir,i} \cdot \omega_{ac,i} \cdot s_i(t))^2 \right)} = \frac{\sum_{i=1}^I \left((\omega_{\tau,i} \cdot \omega_{dir,i} \cdot \omega_{ac,i} \cdot s(t) \cdot g_s \cdot g_i)^2 \cdot \vec{u}_i \right)}{\sum_{i=1}^I \left((\omega_{\tau,i} \cdot \omega_{dir,i} \cdot \omega_{ac,i} \cdot s(t) \cdot g_s \cdot g_i)^2 \right)} = \vec{r}_E \quad (11)$$

Pour une source sonore unique s il est donc possible dans une configuration de haut-parleurs donnée de retrouver sa direction, et donc son angle θ_s par rapport au centre du panorama que l'on situe par convention sur l'axe des ordonnées.

$$\theta_s = \tan^{-1} \frac{y_{r_E}}{x_{r_E}} \quad (12)$$

Il est à noter que l'angle peut être approximé par toute fonction optimisée de la fonction arctangeante ou atan2 (développement polynomial...) pour en réduire le coût de calcul comme par exemple :

$$\tan^{-1}(z) \approx z * \left(\frac{\pi}{4} + 0.273 * (1 - z) \right) \quad (13)$$

On peut déduire l'indice de panning p_s de la source en normalisant l'angle trouvé θ_s par l'ouverture maximum Ω_{max} du panorama sonore de la configuration de haut parleur originale

$$p_s = \frac{\theta_s}{\Omega} \quad (14)$$

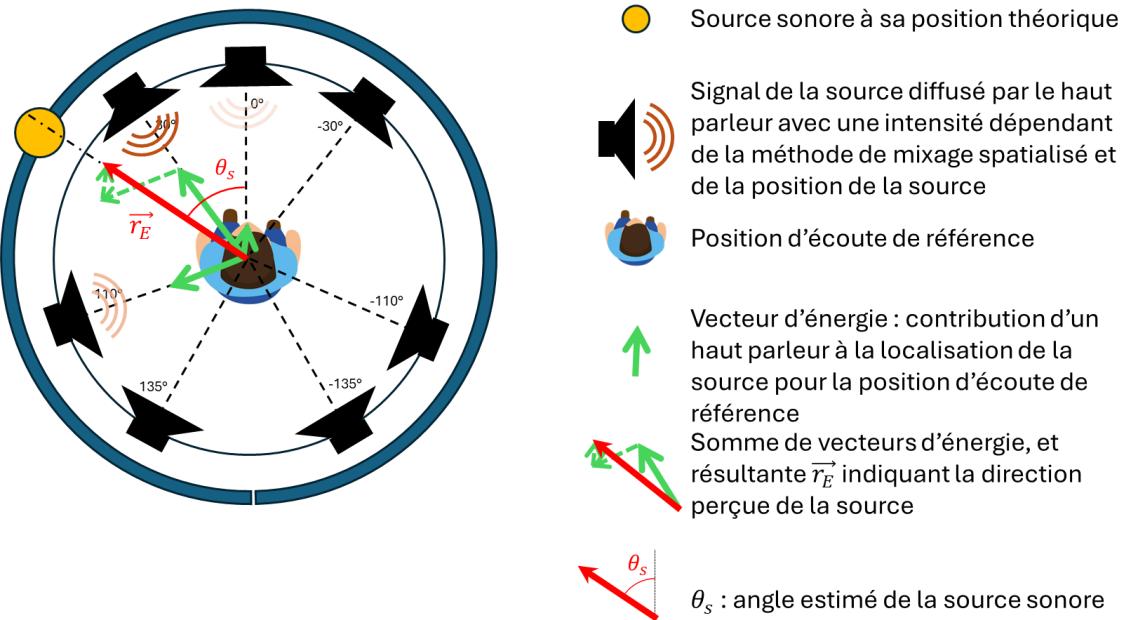


Figure 5: estimation de la localisation d'une source sonore sur un setup multicanal grâce au modèle par vecteur d'énergie

Nous pouvons faire les hypothèses suivantes pour simplifier le modèle (mais le modèle complet peut tout aussi bien être conservé) :

- La position de référence est dans l'axe de chaque haut parleur
 - o pas d'atténuation due à la directivité du haut parleur, donc $\forall i, \omega_{dir,i} = 1$
- Tous les haut-parleurs sont à la même distance de la position d'écoute
 - o pas de différence d'atténuation dû à la propagation acoustique donc $\forall i, \omega_{ac,i} = \omega_{ac}$ constant
- Aucun délai n'est utilisé dans l'algorithme de spatialisation du signal multicanal original
 - o pas de différence de temps d'arrivée donc $\forall i, \Delta t_i = 0 \Rightarrow \forall i, \omega_{\tau,i} = 1$

On a donc en simplifiant l'équation 11 :

$$\overrightarrow{r_E(t)} = \frac{\sum_{i=1}^I (s_i^2(t) \cdot \vec{u}_i)}{\sum_{i=1}^I (s_i^2(t))} \quad (15)$$

Principe de la transposition du panorama

L'idée générale de l'invention est de pouvoir transposer un panorama mixé sur une configuration de haut-parleurs donnée (stéréo ou 5.1 par exemple) vers une nouvelle configuration multicanale. Cette transposition se fait en deux étapes:

- Une extraction de K signaux sonores dans le panorama d'origine correspondant à des portions de panorama
- Une restitution de ces K signaux à une place équivalente dans le nouveau panorama grâce une méthode de spatialisation.

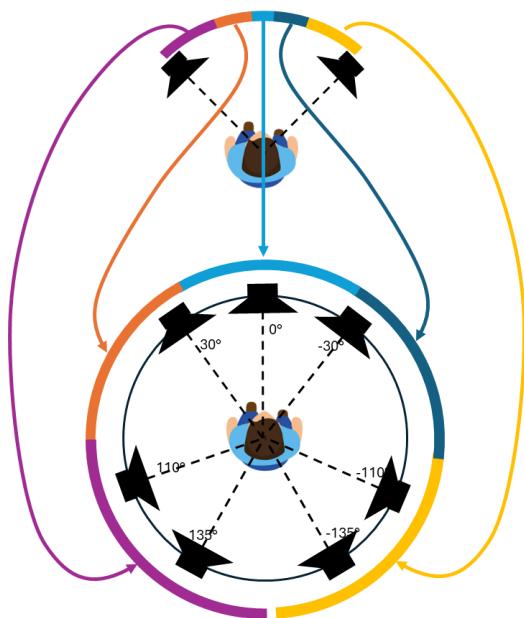


Figure 6: transposition de panorama par extraction et respatialisation

Il est à noter que cette méthode fonctionne aussi bien en upmix (nombre de canaux audio supérieur dans la configuration de haut-parleurs de destination) qu'en downmix (nombre de canaux audio inférieurs dans la configuration de destination).

Pré-traitements

Egalisation

Une première étape (facultative) d'égalisation du signal multicanal entrant permet de réaliser une calibration du système en modifiant la balance tonale du signal pour l'adapter à la configuration de haut-parleur finale et atteindre une cible de réponse en fréquence de calibration. Le filtre d'égalisation est appliqué de manière identique à tous

les signaux du signal multicanal considéré.

Cette étape de calibration peut se trouver en amont (premier traitement) de la chaîne, ou en aval (dernier traitement). On préfèrera pour des raisons de coût de calcul positionner l'égalisation en amont si le nombre de canaux de la configuration initiale est plus faible que le nombre de canaux de la configuration finale, et inversement.

Filtre de re-corrélation des basses fréquences.

Le signal multicanal entrant peut également passer par un filtre de re-corrélation sur des basses fréquences l'ensemble des canaux pour éviter toute fluctuation indésirable dues à l'étape d'extraction spatiale et à l'analyse en fréquence.

Si l'extraction spatiale se fait à l'aide d'une STFT, la précision dans les basses fréquences est moindre que dans les hautes fréquences, et les fluctuations en temps réel du masque de gain d'extraction (voir figure 9) sont caractéristiques d'un traitement non linéaire qui peut introduire une modulation d'amplitude créant des composantes fréquentielles indésirables, surtout pour les fréquences dont la période est plus grande que la fenêtre d'analyse.

Si des composantes basse fréquences décorrélées (décalage temporel entre les canaux, réverbération...) sont présentes dans le signal multicanal d'entrée, cela se traduit dans la phase d'analyse par une oscillation de l'indice de panning estimé, et donc du gain appliqué à cette fréquence qui en dépend directement. En appliquant une re-corrélation (signal identique dans tous les canaux d'entrée), on supprime à ces fréquences cette oscillation et donc le phénomène de distorsion.

La re-corrélation se fait de la manière suivante :

- Un filtre passe bas avec une fréquence de coupure f_{xover1} est appliqué à tous les signaux du signal multicanal entrant.
- Un filtre passe haut complémentaire avec une fréquence de coupure f_{xover2} est appliqué à tous les signaux du signal multicanal entrant.
- L'ensemble des signaux résultants du filtrage passe bas sont sommés pour obtenir un signal mono.
- Ce signal mono est multiplié par un facteur g_{correl} . On peut prendre par défaut $g_{correl} = \frac{1}{\sqrt{I}}$ pour maintenir le niveau RMS du signal global.
- Ce signal mono est ajouté à chacune des composantes du signal multicanal résultant du filtrage passe haut.

Les filtres passe-haut et passe bas constituent un crossover.

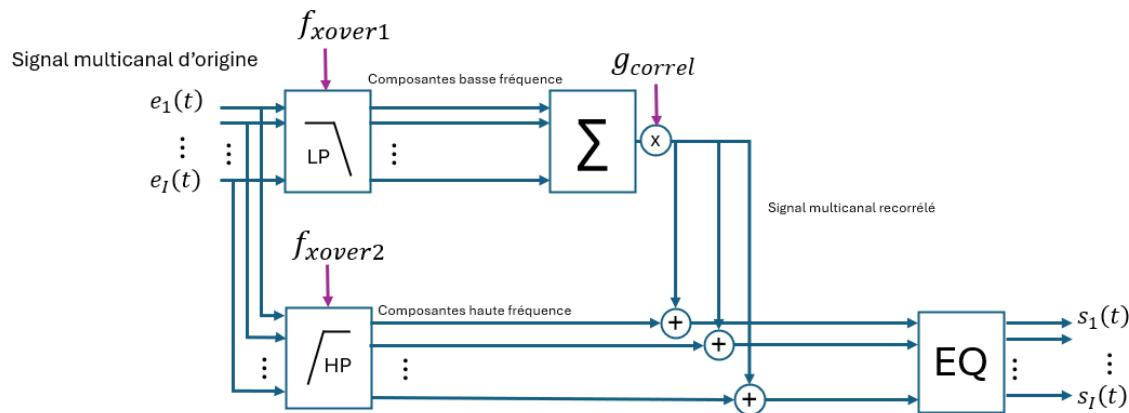


Figure 7: étage de prétraitement incluant le filtre de recorrelation des basses fréquences et l'étage d'égalisation.

Extraction spatiale

L'étage d'extraction permet de générer un signal contenant l'ensemble des composantes sonores spatialisées autour d'un angle donné dans la configuration d'origine. Le traitement se fait dans le domaine fréquentiel afin de bénéficier d'une meilleure séparation des éléments du mix. Voici les étapes de séparation.

1. Transformée de fourier à court terme (STFT, ou tout autre analyse temps fréquence utilisable en temps réel) pour chaque canal d'entrée.
2. Pour chaque fréquence et à chaque instant, estimation de la direction perçue normalisée $p(f, T)$ à l'aide de la méthode précédente

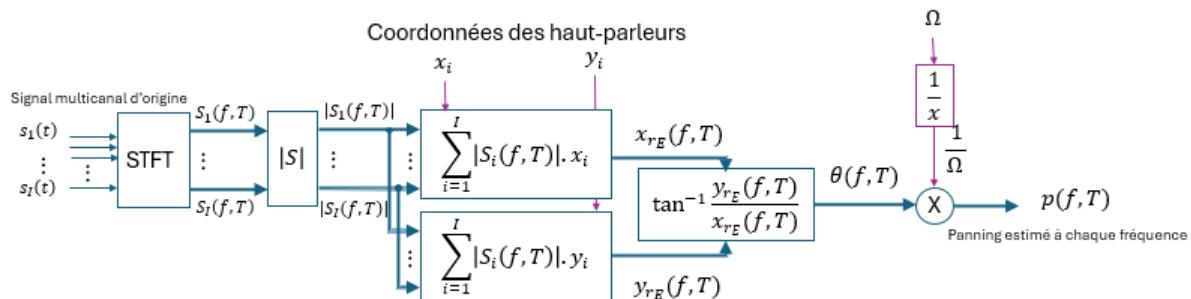


Figure 8 : estimation de l'indice de panning pour chaque fréquence à un instant T pour un signal multicanal donné

3. Pour chaque portion de panorama, définition d'une loi de gain d'extraction dépendant de l'angle de la direction perçue, valant 1 autour de l'angle cible d'extraction et décroissant jusqu'à 0 lorsqu'on s'éloigne de cette valeur cible

soit les constantes suivantes :

- pan_k : l'indice de panning central de la portion de panorama.

$$pan_k \in [-1, 1]$$

- w_k : la largeur de la portion de panorama
- $w_k \in [0, 2 \cdot (1 - |pan_k|)]$
- $slope_k$: la pente du filtre d'extraction en dB sur un demi panorama complet.
- $floor_k$: l'atténuation maximale du filtre d'extraction

Pour $p \in [-1, 1]$, on définit la fonction gain de la manière suivante :

$$G_{dBk}(p) = slope_k \cdot \left(\frac{w_k}{2} - |p - pan_k| \right) \quad (16)$$

G_{dBk} est ensuite seuillé entre $floor_k$ et 0 dB

$$G_{dBk}(p) = \max \left(\min \left(slope_k \cdot \left(\frac{w_k}{2} - |p - pan_k| \right), floor_k \right), 0 \right) \quad (17)$$

On passe ensuite le gain en linéaire :

$$G_k(p) = 10^{\frac{G_{dBk}(p)}{20}} \quad (18)$$

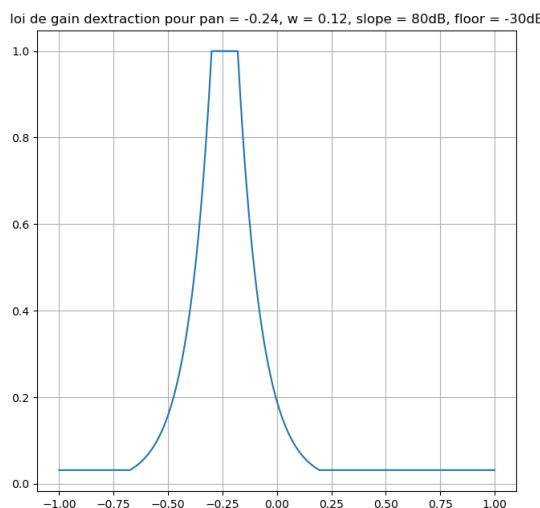


Figure 9: exemple de loi de gain d'extraction

4. Pour chaque fréquence, le gain d'extraction est déduit à l'aide de cette loi en fonction de l'estimation de la direction perçue normalisée $p(f, T)$

$$G_k(f, T) = G_k(p(f, T)) \quad (19)$$

Ce gain peut être soit calculé en temps réel ou pour réduire le coût de calcul, on peut pré-calculer pour chaque portion de panorama k une table de correspondance (lookup table) ou l'on stocke le gain pour chaque valeur de p .

5. Pour chaque fréquence, le signal $G_k(f, T)$ est lissé temporellement à l'aide d'un suivi d'enveloppe et d'un temps de relâchement.

Soit τ_f : le temps de relâchement en secondes pour la fréquence f , on définit :

$$\beta_f = e^{-\frac{N_{hop}}{F_s \tau_f}} \quad (20)$$

Le lissage est appliqué comme suit :

$$G_{smooth_k}(f, T) = \max(G_k(f, T), \beta_f G_{smooth_k}(f, T - 1)) \quad (21)$$

Autre possibilité : lissage rampe plutôt que exponentiel :

$$\Delta_f = \frac{-N_{hop}}{F_s \tau_f} \quad (20 \text{ bis})$$

$$G_{smooth_k}(f, T) = \max(G_k(f, T), G_{smooth_k}(f, T - 1) - \Delta_f) \quad (21 \text{ bis})$$

6. Lissage du gain sur l'axe fréquentiel

Le gain est également lissé sur l'axe fréquentiel par exemple par une convolutions par une fenêtre glissante sur l'axe des fréquences

Soit M la largeur de la fenêtre fréquentielle, W la fenêtre de lissage

$$\widetilde{G}_k(f, T) = \frac{\sum_{d=-D}^D W(d) \widetilde{G}_k(f-d, T)}{\sum_{d=-D}^D W(d)}, \text{ avec } D = \frac{M-1}{2} \quad (22)$$

Remarque : si la fenêtre W est rectangulaire (toutes les valeurs à 1) on obtient un lissage par moyenne glissante.

7. Application du gain au signal du haut parleur le plus proche de l'angle cible

Soit h_k l'indice du haut parleur le plus proche de l'angle défini comme $\vartheta_k = pan_k \Omega_{max}$

Dans le domaine fréquentiel, on définit le signal $Q_k(f, T)$:

$$Q_k(f, T) = S_{h_k}(f, T) \widetilde{G}_k(f, T) \quad (23)$$

8. Transformée de fourier à court terme inverse (ou méthode inverse correspondant à l'analyse temps fréquence utilisée) pour resynthétiser le signal extrait $q_k(t)$ correspondant à la portion de panorama k dans le domaine temporel

$$q_k(t) = \text{iSTFT}(Q_k(f, T)) \quad (24)$$

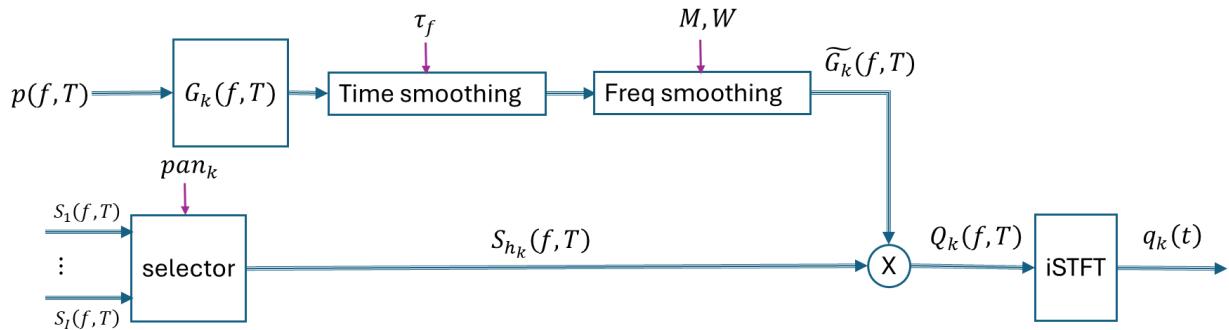


Figure 10: génération et application du masque de sélection

TDAP - Triangular Distribution Amplitude Panning (avec correction itérative par vecteur d'énergie)

Contexte & objectifs

Cette nouvelle méthode de panning permet de positionner une source virtuelle à un angle cible sur un réseau de haut-parleurs au moyen d'une distribution triangulaire de gains, tout en garantissant un niveau global constant et une cohérence directionnelle (via une correction itérative).

Avantages par rapport aux méthodes existantes

Cette méthode assure au sweet spot une cohérence directionnelle parfaite, un réglage simple de la largeur de source et un calcul peu coûteux.

Hypothèses & conventions

Les angles sont exprimés en degrés.

Le sweet spot du setup de haut-parleur (point d'écoute privilégié) est considéré comme étant l'origine du repère cartésien.

Variables

- **Entrées :**
 - θ_{pan} : angle de panoramique demandé.

- W : largeur triangulaire.
- $\{l_j\} = \{(x_j, y_j)\}$: positions des haut-parleurs (tableau de taille $[J, 2]$).
- α_{min} (optionnel) : borne inférieure de l'intervalle de normalisation de l'angle par défaut, $\alpha_{min} = \theta_{pan} - 180^\circ$
- **Sorties :**
 - $G = \{g_j\}$ (tableau de taille J , linéaire) : gains par haut-parleur.
 - \vec{r}_E : vecteur d'énergie résultant (estimation directionnelle).
- **Autres variables :**
 - α_j : azimut du haut parleur j.
 - $\Delta_j = |\alpha_j - \theta_{pan}|$: distance angulaire locale.
 - δ : erreur angulaire entre \vec{r}_E et θ_{pan} .
 - μ : pas de correction (par défaut 0.8).
 - tol : tolérance angulaire (par défaut 1°)
 - $N_{iter_{max}}$: nombre d'itérations maximal de la boucle d'ajustement (100 par défaut)

1. Pré-traitements

- W peut être limité à une plage $[0 - 300^\circ]$
- Calcul des azimuts haut-parleurs :

$$\alpha_j = \tan^{-1}\left(\frac{y_j}{x_j}\right) \quad (25)$$

- Normalisation de l'angle de panning entre 0 et 360° :

$$\theta_{pan} = \theta_{pan} \bmod 360^\circ \quad (26)$$

- Normalisation des azimuths des haut-parleurs dans l'intervalle $[\alpha_{min}, \alpha_{min} + 360^\circ]$:

$$\forall j, \alpha_j = ((\alpha_j - \alpha_{min}) \bmod 360^\circ) + \alpha_{min} \quad (27)$$

2. Génération des gains (profil triangulaire)

- **Largeur minimale de sécurité** : on souhaite qu'au moins deux haut-parleurs aient un gain non nul.

$$W \leftarrow \max(W, \text{second_smallest}(|\alpha_j - \theta_{pan}|)) \quad (28)$$

- **Calcul de la distribution triangulaire des gains :**

- Calcul des gains initiaux :

$$\forall j, g_j = \max\left(\frac{W - |\alpha_j - \theta_{pan}|}{W}, 0\right) \quad (29)$$

- Normalisation des gains :

$$\forall j, g_j = \frac{g_j}{\sqrt{\sum_j g_j^2}} \quad (30)$$

3. Estimation directionnelle (vecteur d'énergie)

- Calcul du vecteur d'énergie :

$$\overrightarrow{r_E(t)} = \frac{\sum_{i=1}^I (g_i^2(t) \cdot \vec{u}_i)}{\sum_{i=1}^I (g_i^2(t))} \quad (31)$$

- Estimation de l'angle perçu de la source :

$$\theta_s = \tan^{-1} \frac{y_{r_E}}{x_{r_E}} \quad (32)$$

- Calcul de l'erreur :

$$\delta = \theta_s - \theta_{pan} \quad (33)$$

4. Boucle d'ajustement (contrôle itératif)

Le calcul des gains peut donner une direction perçue de la source sonore décalée de la direction voulue, on procède donc par ajustements successifs de l'angle de panning pour minimiser cette erreur.

- Critère d'arrêt : $|\delta| \leq tol$ ou nombre d'itération $> N_{iter_{max}}$
- Mise à jour des gains si le critère d'arrêt n'est pas atteint :

$$\theta_{pan} = \theta_{pan} + \mu \cdot \delta \quad (34)$$

On calcule à nouveau les gains avec la même méthode en répétant les étapes 2 3 et 4 jusqu'à ce que le critère d'arrêt soit vrai.

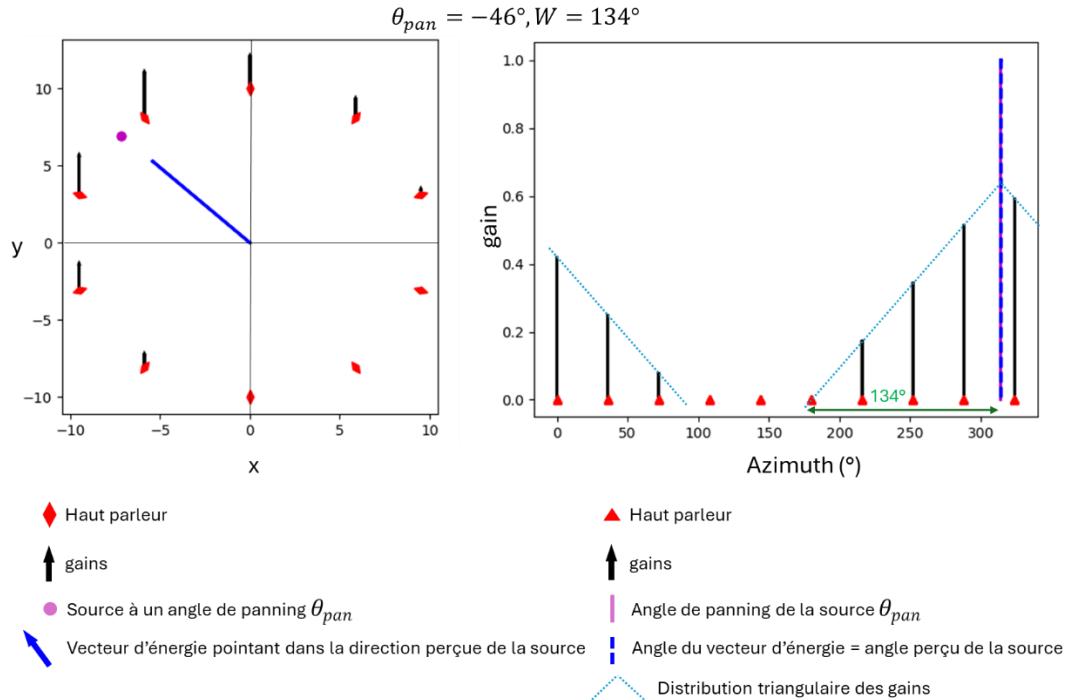


Figure 11: illustration de la méthode TDAP sur un ensemble de 10 haut-parleurs

5. Solution analytique

La boucle d'ajustement peut être remplacée par une résolution analytique du problème.

Pour un θ donné (centre du triangle), le poids **avant normalisation** d'un HP j actif est

$$t_j(\theta) = \frac{W - |\alpha_j - \theta|}{W} \geq 0, \text{ et } w_j(\theta) = t_j(\theta)^2 \quad (35)$$

Sur un intervalle où le signe $\sigma_i = \text{sign}(\alpha_j - \theta) \in \{-1, +1\}$ est constant et où j reste actif ($|\alpha_j - \theta| < W$),

$$t_j(\theta) = \frac{W - \sigma_j(\alpha_j - \theta)}{W} = \frac{c_j + d_j\theta}{W} \quad (36)$$

Avec $c_j = W - \sigma_j\alpha_j$ et $d_j = \sigma_j$

Donc

$$w_j(\theta) = \left(\frac{c_j + d_j\theta}{W} \right)^2 = \frac{a_j\theta^2 + b_j\theta + c'_j}{W^2} \quad (37)$$

Avec $a_j = d_j^2 = 1$, $b_j = 2c_jd_j$, et $c'_j = c_j^2$

Le vecteur d'énergie normalisé s'écrit alors

$$\vec{r}_E(\theta) = \sum_{i \in \mathcal{A}} w_i(\theta) \vec{u}_i = \frac{1}{W^2} (\vec{A}\theta^2 + \vec{B}\theta + \vec{C}) \quad (38)$$

Où \mathcal{A} est l'ensemble actif sur l'intervalle considéré, et

$$\vec{A} = \sum_{i \in \mathcal{A}} a_i \vec{u}_i, \quad \vec{B} = \sum_{i \in \mathcal{A}} b_i \vec{u}_i, \quad \text{et} \quad \vec{C} = \sum_{i \in \mathcal{A}} c_i' \vec{u}_i \quad (39)$$

Pour aligner $r_E(\theta)$ avec la direction cible $\vec{u} = \begin{bmatrix} \cos \theta_{cible} \\ \sin \theta_{cible} \end{bmatrix}$, on impose l'orthogonalité à $\vec{u}_\perp = \begin{bmatrix} -\sin \theta_{cible} \\ \cos \theta_{cible} \end{bmatrix}$:

$$\vec{u}_\perp \cdot \vec{r}_E(\theta) = 0 \quad (40)$$

Et donc

$$\frac{1}{W^2} (\vec{u}_\perp \cdot \vec{A}\theta^2 + \vec{u}_\perp \cdot \vec{B}\theta + \vec{u}_\perp \cdot \vec{C}) = 0 \quad (41)$$

En posant

$$d = \vec{u}_\perp \cdot \vec{A}, e = \vec{u}_\perp \cdot \vec{B}, \text{ et } f = \vec{u}_\perp \cdot \vec{C}$$

On doit chercher les racines d'un polynôme du second degré dont la solution est :

$$\theta^* = \frac{-e \pm \sqrt{e^2 - 4df}}{2d} \quad (42)$$

On garde la racine qui tombe dans l'intervalle courant où \mathcal{A} et σ_i sont constants

Re-spatialisation

Pour restituer la scène sonore sur le système électroacoustique de destination, il est nécessaire de spatialiser les K signaux générés à l'aide de l'étage d'extraction sur les J haut-parleurs de la configuration finale.

On définit pour ça :

- Une position d'écoute centrale notée $PE_c = \begin{bmatrix} x_{PE_c} \\ y_{PE_c} \end{bmatrix}$
- Un angle d'ouverture du nouveau panorama sonore noté Φ
- Un angle correspondant à l'orientation du centre du nouveau panorama noté φ_c
- K angles de spatialisation notés ψ_k pour chacun des signaux $q_k(t)$ extraits à des indices de panning pan_k tel que :

$$\forall k, \psi_k = \varphi_c + \frac{\Phi}{2} pan_k \quad (43)$$

- Un facteur d'élargissement notée $spread$

- K angles de largeur de spatialisation notés χ_k pour chacun des signaux $q_k(t)$ extrait avec une largeur w_k tel que :

$$\forall k, \chi_k = w_k \cdot \frac{\Phi}{2} \cdot spread \quad (44)$$

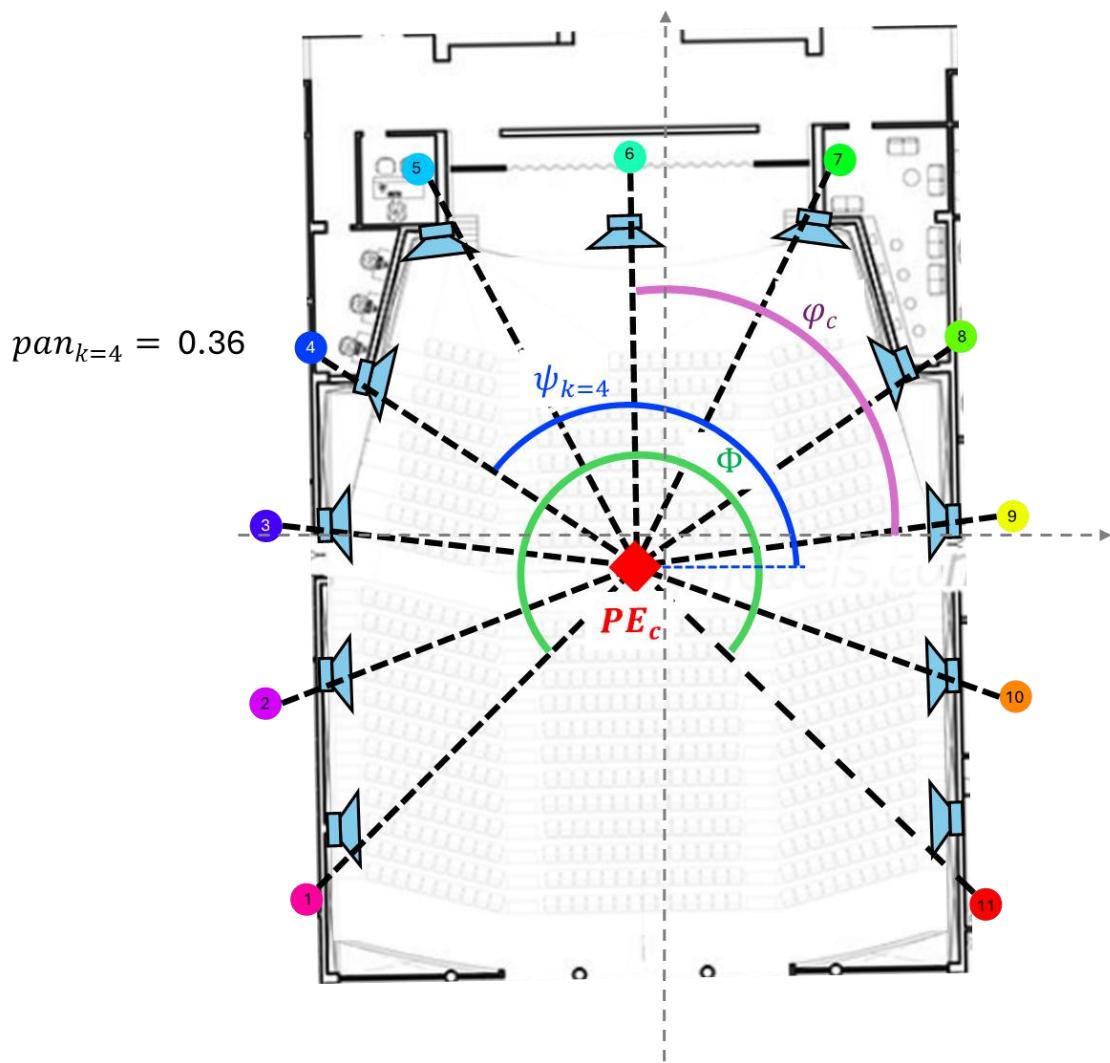


Figure 12: placement de sources sonores en fonction de l'ouverture et de l'orientation de la scène sonore souhaitée

Pour spatialiser le signal $q_k(t)$ sur l'ensemble des haut-parleurs de la configuration finale, on utilise la méthode TDAP précédemment décrite (équations 25 à 34) avec les entrées suivantes :

- ψ_k : angle de panoramique demandé.
- χ_k : largeur triangulaire.
- $L = \{l_j - PE_c\} = \{(x_j - x_{PE_c}, y_j - y_{PE_c})\}$: positions des haut-parleurs relatives au point d'écoute central

$$\circ \quad \alpha_{min} = \varphi_c - 180^\circ$$

Pour chaque signal $q_k(t)$, on obtient en sortie un jeu de gains pour les haut-parleurs que l'on note $g_{k,j}$.

Le signal de chaque haut-parleur de la configuration finale est construit de la manière suivante :

$$\forall j, r_j(t) = \sum_{k=1}^K g_{k,j} \cdot q_k(t) \quad (45)$$

Références

Bibliographie :

- [1] E. Kurz, « Efficient prediction of the listening area for plausible reproduction », institut für elektronische musik und akustik, Graz, Austria, 2018.

Brevets concurrents : à completer.