Numéros / n° 7-8 - Culture du code

Â«Â Une approche perceptive pour la spatialisation du sonÂ Â»

Sylvain Marchand

Résumé

Cet article est la trace écrite de la conférence invitée donnée lors des Journées d?informatique musicale 2019 qui ont eu lieu du 13 au 15 mai 2019 à Bayonne. Il dresse un rapide historique de la mise en espace en musique, donne quelques notions de physique pour la propagation du son, décrit des mécanismes du système auditif humain pour la localisation des sources sonores, puis passe en revue les méthodes de spatialisation actuelles en listant leurs avantages et inconvénients. Pour finir, une nouvelle approche, basée sur la perception, est présentée.

Introduction

En musique, certaines caractéristiques revêtent une importance particulière, comme la hauteur, la durée, l?intensité et le timbre, mais aussi la mise en espace?

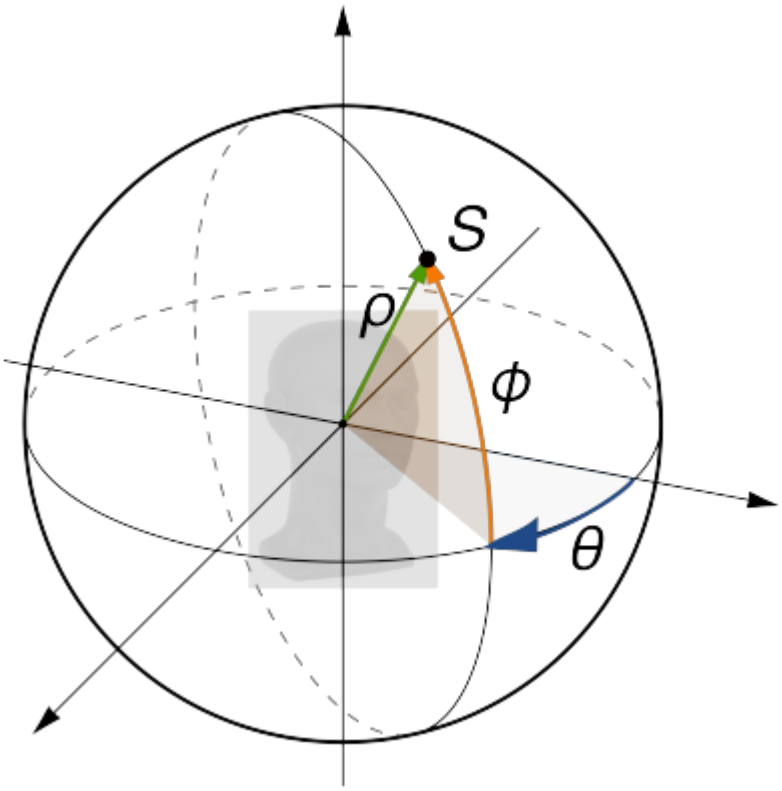
Quand on pense à l?utilisation de la mise en espace dans la musique, on pense tout de suite aux développements techniques du xx^{e} siècle, comme l?électronique et l?informatique, et à la musique électroacoustique. Cependant, la spatialisation en musique est beaucoup plus ancienne. Elle remonte même à l?Antiquité, puisque l?antiphonie, au iv^{e} siècle, consistait à faire dialoguer deux ensembles vocaux situés à des endroits différents. Notons aussi Andrea Gabrieli (xvi^{e} siècle), mettant en scène deux ensembles (orgues et ch?ur) dans la basilique Saint-Marc à Venise. On peut citer Jean-Sébastien Bach, qui, dans la *Passion selon Saint Matthieu* (1729), place 3 ch?urs à différents endroits. Néanmoins, les projets techniques ont en effet permis une évolution rapide au xx^{e} siècle. Walt Disney déploie un système de 33 microphones et 90 haut-parleurs pour *Fantasia* (1939). Le record est battu par Edgard Varèse et son *Poème électro*

Si l?on s?intéresse à la physique du son, il nous faut parler d?onde et de trajet acoustiques. L?onde de pression acoustique se déplace dans l?air à la vitesse d?environ 340 mètres par seconde (c : célérité du son), ce qui forme un ou plusieurs trajet(s) acoustique(s) entre la source et le récepteur. Il y a en effet une onde directe, parvenant de la source au récepteur sans obstacle, qui est suivie d?ondes réfléchies arrivant plus tard et atténuées, du fait des obstacles rencontrés qui se comportent comme des miroirs acoustiques. Ceci forme une réponse impulsionnelle, utilisée par exemple pour reproduire l?effet de réverbération. Plus précisément, si s est le son de la source et h la réponse impulsionnelle de la salle, alors $s * h$ (où $*$ désigne la convolution) est le son résultant, c?est-à-dire la source jouant dans la salle. Si on passe dans le domaine spectral, en utilisant la transformée de Fourier pour obtenir S et H, spectres respectifs de s et h (on dit aussi que S est le spectre du signal s et que H est la fonction de transfert de la réponse impulsionnelle h), alors le résultat est $S \times H$ (la multiplication remplaçant la convolution). Les choses se compliquent encore si l?auditeur ou la source est en mouvement. Par exemple, si l?on considère une source émettant une sinusoïde de fréquence F, on mesure bien cette fréquence au niveau de l?auditeur quand la vitesse relative entre l?auditeur et la source est nulle. Cependant, si l?auditeur se rapproche de la source à la vitesse du son, c?est une sinusoïde de fréquence $2F$ qui sera perçue (effet Doppler). Si c?est la source qui s?approche de l?auditeur à la vitesse du son, on observe une accumulation d?énergie (le mur du son).

Dans cet article, on se placera dans des conditions simplifiées : en champ libre (sans obstacle), avec des

sources sonores supposées fixes, ponctuelles et omnidirectionnelles. Les positions seront décrites en coordonnées sphériques centrées sur la tête de l?auditeur (distance ?, azimut ?, élévation ?), voir Figure 1.

Figure 1. Coordonnées sphériques centrées sur la tête de l?auditeur



Source : Sylvain Marchand, 2020

Les problématiques abordées seront alors :

- la localisation : à partir d?enregistrements pris à des positions déterminées (par exemple au niveau des deux oreilles), on tente de retrouver la position de la source sonore (applications : suivi de trajectoires, séparation de sources, etc.) ;
- la spatialisation : à partir d?un son et d?une position dans l?espace, on veut faire croire que le son est émis depuis cette position (applications : musique électroacoustique, réalité virtuelle, etc.).

1. Localisation sonore

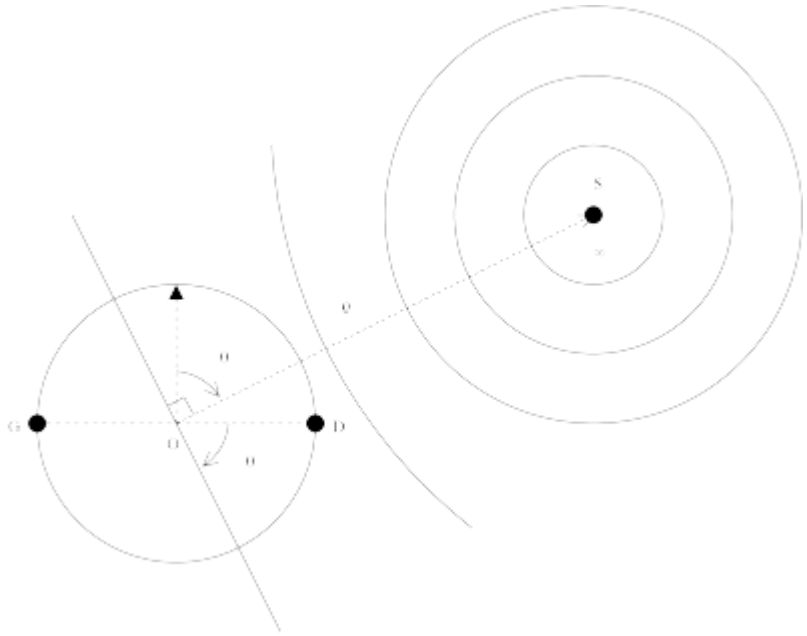
Pour localiser une source sonore, le système auditif humain utilise des indices acoustiques (Blauert, 1997). Si l?on se limite au cas de l?azimut de la source, ce sont les indices interauraux qui prévalent. Il existe aussi des indices monauraux (notamment pour la perception de l?élévation), mais aussi des indices dynamiques, etc.

1.1. Indices interauraux

Quand une onde sonore émise depuis une source arrive au niveau de la tête de l?auditeur, elle arrive à des instants séparés et avec des intensités légèrement différentes entre les deux oreilles, car elle ne suit pas

exactement les mêmes trajets (voir Figure 2).

Figure 2. Onde sonore émise par la source et trajets vers les oreilles de l?auditeur



Source : Sylvain Marchand, 2019

Dans le plan, la source est localisée par son azimut *?* et sa distance *?*. On désigne par *r* le rayon de la tête de l?auditeur et on suppose que *?* est très supérieur à *r*. En fait, la source *S* est supposée être à l?infini, de sorte que l?onde arrivant à l?auditeur peut être supposée plane.

Si on s?intéresse aux trajets entre la source *S* et les oreilles gauche (*G*) et droite (*D*) de l?auditeur, supposé placé au centre du repère *O*, on a :

$$\rho_G = \rho + \Delta_\rho$$
$$\rho_D = \rho + \Delta_\rho$$

avec

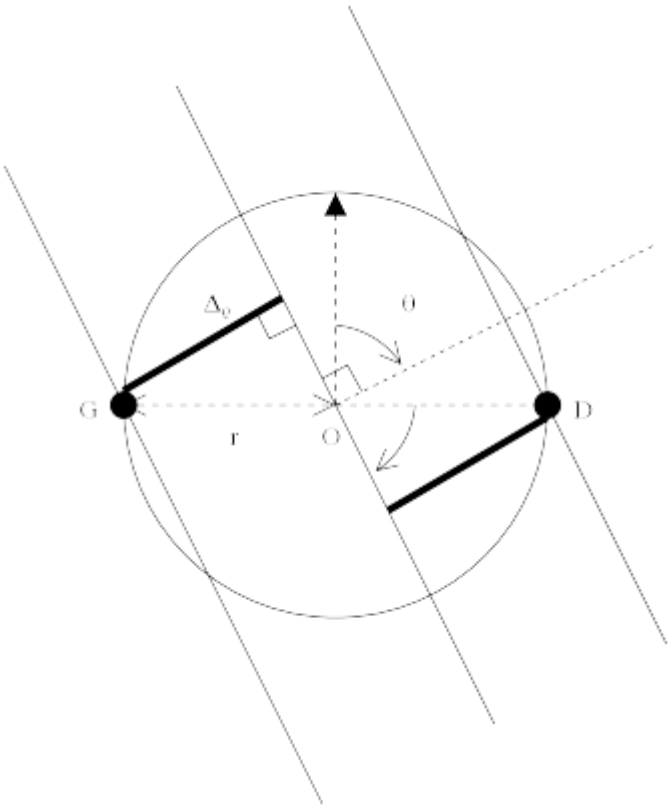
$$\Delta_\rho = r\sin(\theta)$$

et par conséquent, la différence de temps entre les oreilles est :

$$\Delta_T - \frac{\rho_G = \rho_D}{c} = \frac{2\Delta_\rho}{c} = 2r/c\sin(\theta)$$

où *c* est la célérité du son. Cela correspond aux travaux d?Eric von Hornbostel et Max Wertheimer (1920), qui négligeaient l?effet de la tête de l?auditeur (Figure 3).

Figure 3. Propagation de l?onde sonore en négligeant l?effet de la tête

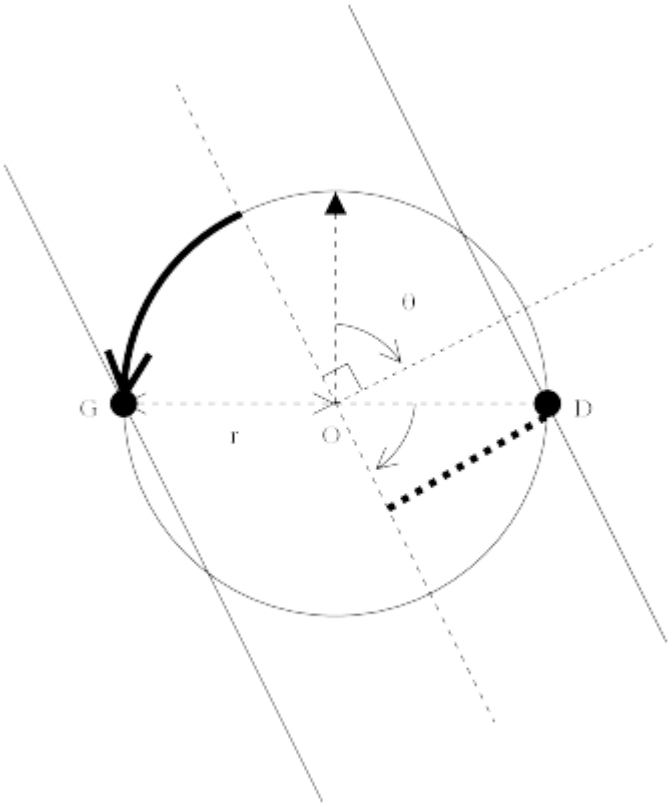


Source : Sylvain Marchand, 2019

En fait, la tête est un obstacle à l?onde acoustique, qui doit la contourner. C?est ce qu?on appelle l?ombre de la tête (Figure 4), qui doit être prise en compte, ce qui donne (Woodworth et Schlosberg, 1954) :

$$\Delta_T(\theta, f) = r/c(\sin(\theta) + \theta)$$

Figure 4. Propagation de l?onde sonore avec prise en compte de l?ombre de la tête



Source : Sylvain Marchand, 2019

En fait, la tête n?est pas ronde et il convient d?appliquer un facteur correctif qui varie en fonction de la fréquence (Wightman et Kistler, 1997). Nous avons montré que ce modèle pouvait être simplifié ainsi (Mouba *et. al.*, 2008), ?$_T$ correspondant en vérité à la différence interaurale de temps (*Interaural Time Difference* ou ITD) :

$$\mathrm{ITD}(\theta, f) = \beta(f) r/c \sin(\theta)$$

En ce qui concerne la différence interaurale d?intensité (*Interaural Level Difference* ou ILD), puisque l?intensité sonore *I* est inversement proportionnelle au carré de la distance, on obtient :

$$\begin{aligned}\Delta_L &= 10\log_{10}(I_D) - 10\log_{10}(I_G) \\ &= C\log(\rho_G/\rho_D) \\ &= C\log\left(\frac{\rho+\Delta_\rho}{\rho-\Delta_\rho}\right) \\ &= C\sum_{n=0}^{\infty}\frac{1}{2n+1}\left(\frac{\Delta_\rho}{\rho}\right)^{2n+1}\end{aligned}$$

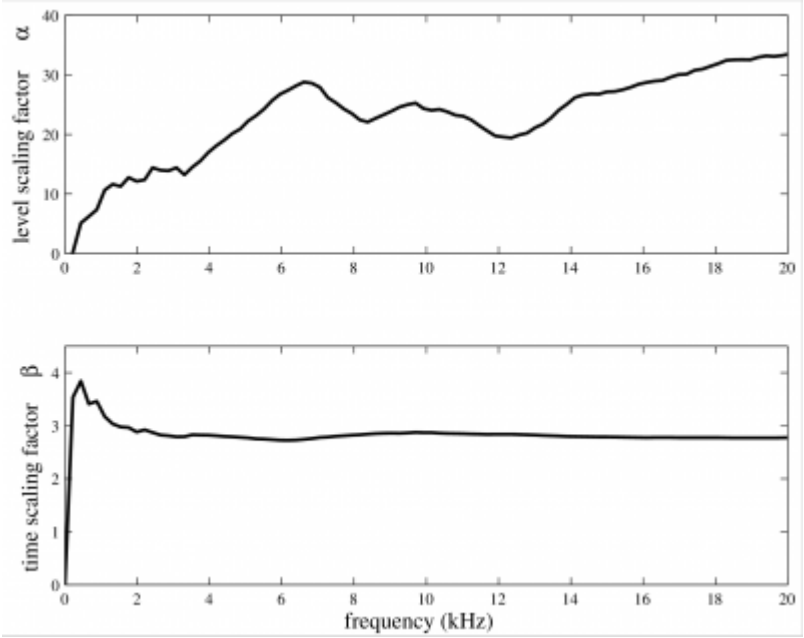
en faisant un développement limité. Pour ?$_?$? *?*, l?ordre 1 est une bonne approximation et l?ILD (en décibels ou dB) est proportionnelle à l?ITD. Finalement, on aboutit à (Viste, 2004) :

$$\mathrm{ILD}(\theta, f) = \alpha(f)/c \sin(\theta)$$

Les facteurs multiplicatifs *?* et *?* (Figure 5) sont obtenus à partir de la base de données du Cipic (Algazi

et. al., 2001), qui recense les chemins acoustiques pour différents auditeurs et différentes positions de la source. Les erreurs moyennes commises par les modèles sont faibles (respectivement 4,29 dB et 0,052 ms pour les ILD et les ITD).

Figure 5. Facteurs multiplicatifs ? (haut) et ? (bas) pour les modèles d?ILD et d?ITD proposés



Source : Sylvain Marchand, 2019

1.2. Chemins acoustiques

En réalité, les chemins acoustiques entre la source et la tête sont des réponses impulsionnelles, appelées en anglais *Head-Related Impulse Responses* (HRIR). C?est ce que contient la base de données du Cipic (voir plus haut). Leurs versions spectrales sont les *Head-Related Transfer Functions* (HRTF). Il s?agit de paires de fonctions (une fonction pour chaque oreille) dépendant de la position de la source sonore, la fréquence du son émis, mais aussi de la morphologie de l?auditeur (oreilles, tête, buste?).

Nous en avons proposé, dans un article (Mouba et Marchand, 2006) une version synthétique respectant les indices interauraux :

$$H_G(\theta,f) = 10^{+\Delta_\alpha(\theta,f)/2_\alpha + i\Delta_\phi(\theta,f)/2}$$
$$H_D(\theta,f) = 10^{-\Delta_\alpha(\theta,f)/2_\alpha + -\Delta_\phi(\theta,f)/2}$$

avec

$$\Delta_\alpha(\theta,f) = \mathrm{ILD}(\theta,f)/20$$
$$\Delta_\phi(\theta,f) = \mathrm{ITD}(\theta,f) \cdot 2\pi f$$

et, comme indiqué dans l?introduction, il est possible d?obtenir un rendu binaural simple à partir du spectre de la source *S* comme suit :

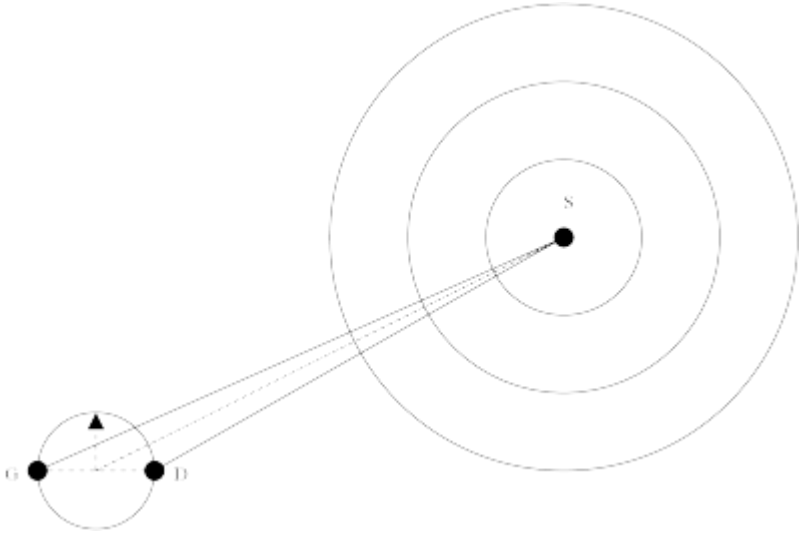
$$S_G(\theta, f) = H_G(\theta, f) \cdot S(f)$$
$$S_D(\theta, f) = H_D(\theta, f) \cdot S(f)$$

même si, en toute rigueur, il s?agit alors de spatialisation sonore.

2. Spatialisation sonore

La spatialisation, telle que définie plus haut, permet de faire croire, à partir d?un son et d?une position dans l?espace, que le son est émis depuis cette position. Ici, il s?agit donc de faire croire à une source dont le son se propagerait en champ libre (Figure 6).

Figure 6. Cas idéal d?une source dont le son se propage en champ libre

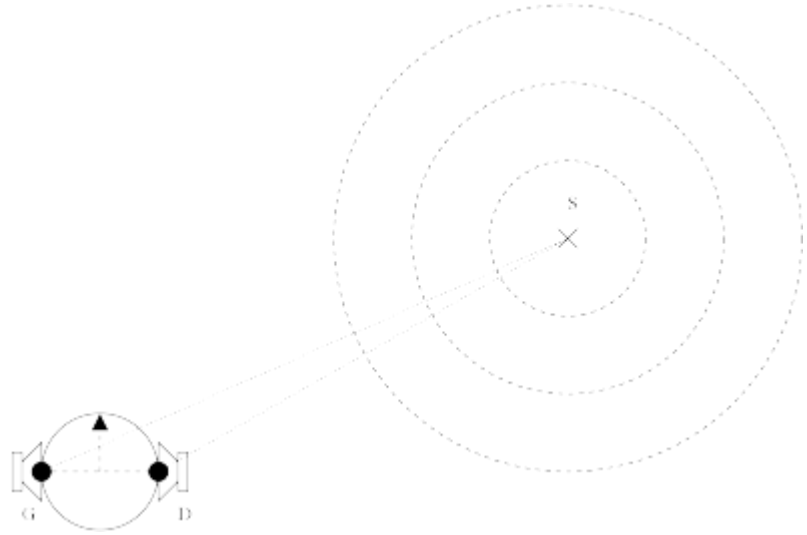


Source : Sylvain Marchand, 2019

2.1. Technique binaurale

La technique binaurale est plus ancienne qu?on ne peut le croire, puisqu?initialement proposée par les laboratoires Bell en 1920. La diffusion se fait à l?aide d?un casque d?écoute (2 petits haut-parleurs, directement au niveau des oreilles), comme on peut le voir dans la Figure 7 et l?enregistrement peut se faire à l?aide de 2 microphones placés à l?intérieur des oreilles. Comme nous l?avons montré plus haut, la synthèse des signaux binauraux est possible (voir équations précédentes, ou au prix du stockage des HRIR en mémoire et de leur interpolation pour les méthodes plus classiques).

Figure 7. Source virtuelle reproduite à l?aide de la technique binaurale



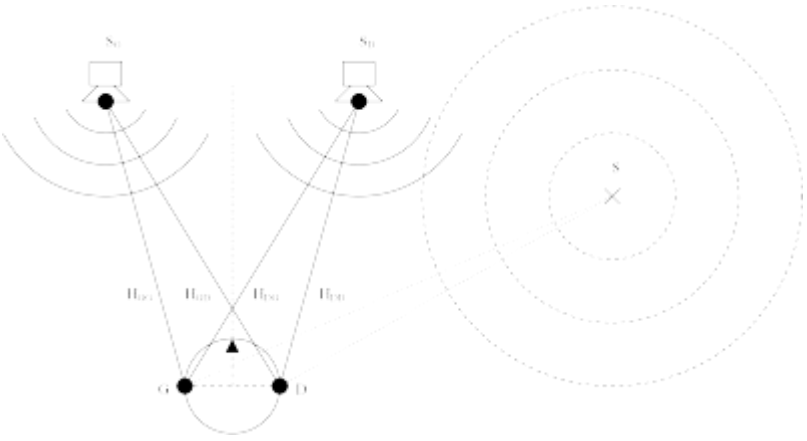
Source : Sylvain Marchand, 2019

2.2. Technique transaurale

Toutefois, une technique nécessitant un casque d?écoute est contraignante, car invasive. Heureusement, il existe des techniques reposant sur des haut-parleurs externes. Cela peut aussi permettre d?envisager une diffusion de musique à destination d?un public d?auditeurs (et pas un seul auditeur). Ainsi, la technique transaurale vise à reproduire les signaux binauraux au niveau des oreilles de l?auditeur, mais à partir d?une diffusion à l?aide de 2 haut-parleurs externes (l?auditeur devant se tenir à égale distance des deux haut-parleurs). Or, avec 2 haut-parleurs et 2 oreilles, il y a 4 trajets acoustiques possibles (Figure 8) et donc 4 fonctions de transfert *H* d?un haut-parleur (*G* ou *D*) vers une oreille (*G* ou *D*). Il s?agit alors de compenser les trajets croisés, d?un haut-parleur vers l?oreille du côté opposé. Cela peut se faire en inversant le système d?équations suivant :

$$\begin{bmatrix} S'_G \\ S'_D \end{bmatrix} = \begin{bmatrix} H_{GG} & H_{DG} \\ H_{GD} & H_{DD} \end{bmatrix} \cdot \begin{bmatrix} S_G \\ S_D \end{bmatrix}$$

Figure 8. Source virtuelle reproduite à l?aide de la technique transaurale



Source : Sylvain Marchand, 2019

L?aspect contraignant de cette technique est l?obligation de calibrer précisément les chemins acoustiques.

2.3. Stéréophonie

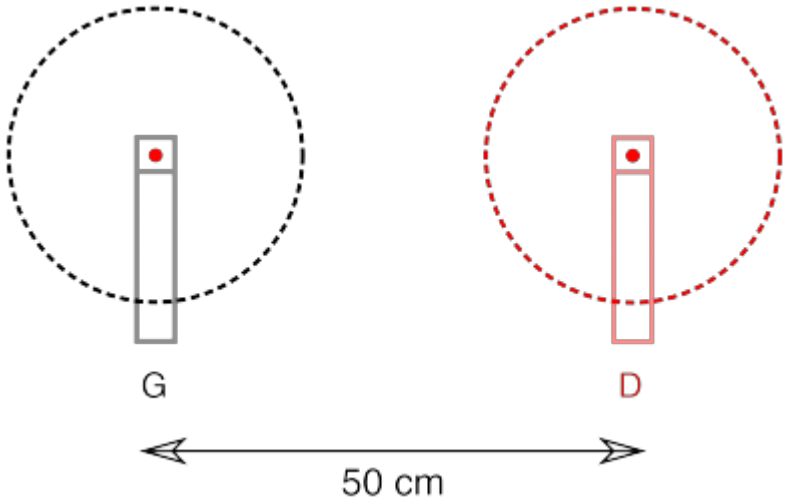
La stéréophonie proposée par Alan Blumlein en 1931 est, de ce point de vue, une technique plus robuste. Il s?agit d?une diffusion à l?aide de 2 haut-parleurs formant un angle de 60 degrés, l?auditeur complétant le triangle équilatéral. Cette technique a connu un essor considérable et est encore prédominante aujourd?hui, même si, en pratique, les contraintes précédentes ne sont pas toujours respectées.

2.3.1. Enregistrement (captation)

Il est possible d?enregistrer des signaux stéréophoniques, toujours avec un couple de microphones, mais avec plusieurs approches.

Le couple A-B (Figure 9) utilise des microphones omnidirectionnels espacés de 50 cm, ce qui privilégie les différences de temps (ITD).

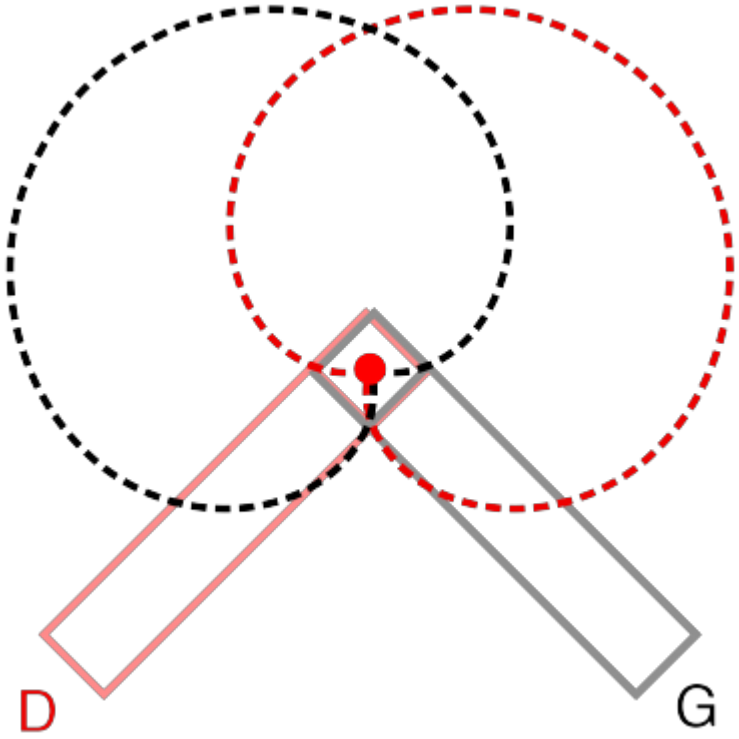
Figure 9. Captation stéréophonique à l?aide d?un couple A-B



Source : Sylvain Marchand, 2020

Le couple X-Y (Figure 10) utilise des microphones cardioïdes de manière coïncidente, ce qui privilégie les différences d?intensité (ILD).

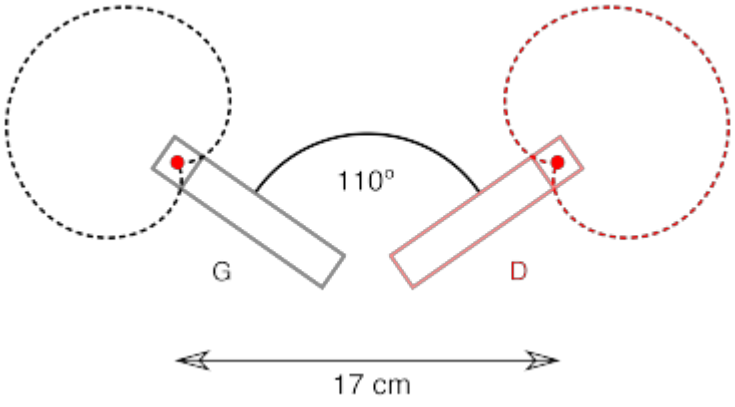
Figure 10. Captation stéréophonique à l?aide d?un couple X-Y



Source : Sylvain Marchand, 2020

Le couple ORTF (utilisé à l?Office de radiodiffusion-télévision française), avec des microphones cardioïdes faisant un angle de 110 degrés et espacés de 17 cm (ce qui correspond approximativement à l?angle et à la distance entre les oreilles), a pour objectif de se rapprocher de la configuration d?une tête humaine pour permettre un compromis entre les deux types d?indices.

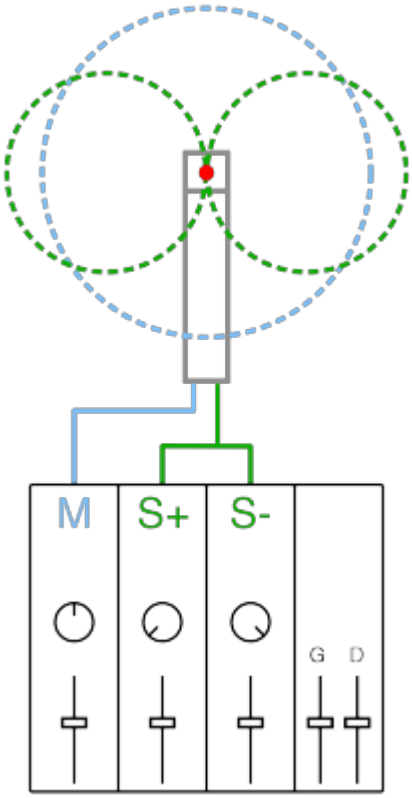
Figure 11. Captation stéréophonique à l?aide d?un couple ORTF



Source : Sylvain Marchand, 2020

Pour finir, le couple M/S (Figure 12) permet une compatibilité monophonique (la composante *M* étant le signal monophonique) et un réglage de l?effet stéréophonique par un ingénieur du son.

Figure 12. Captation stéréophonique à l?aide d?un couple M/S



Source : Sylvain Marchand, 2020

2.3.2. Synthèse (simulation)

Il est également possible de synthétiser en studio des signaux stéréophoniques, par exemple à l?aide de lois de répartition (ou *panning* en anglais). Il en existe plusieurs, comme la loi de répartition à puissance constante :

$$s^2 = s_1^2 + s_2^2$$

avec, par exemple :

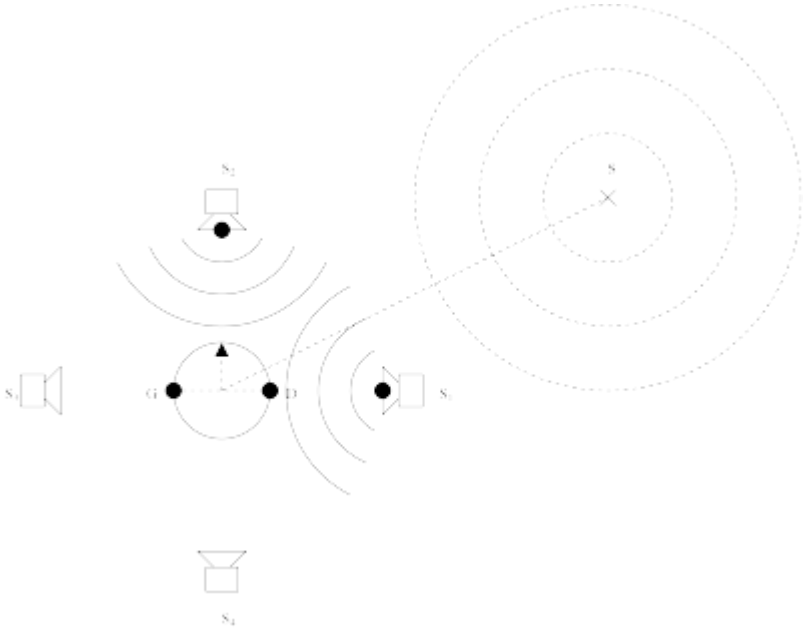
$$s_1(t) = c_1(\theta) \cdot s(t)$$
$$s_2(t) = c_2(\theta) \cdot s(t)$$

avec

$$c_1(\theta) = \sin(\theta)$$
$$c_2(\theta) = \cos(\theta)$$

Cela ouvre aussi la voie à la multidiffusion, avec plus de haut-parleurs et à destination d?un public plus large (Figure 13).

Figure 13. Source virtuelle reproduite à l?aide de plusieurs haut-parleurs



Source : Sylvain Marchand, 2019

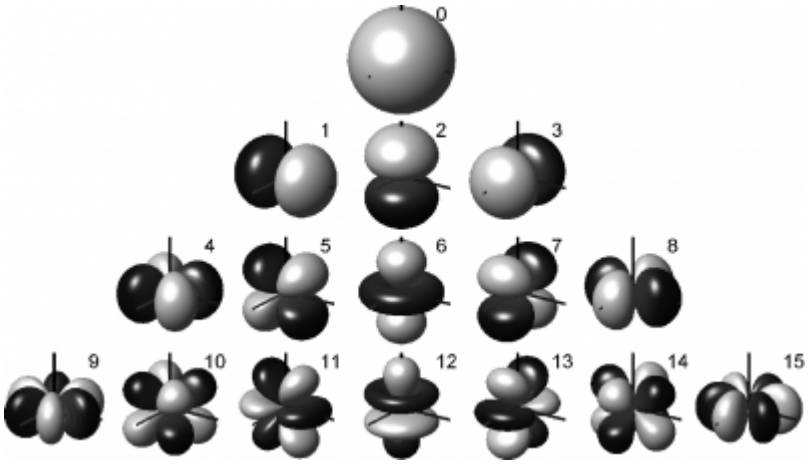
La méthode VBAP (*Vector Base Amplitude Panning*) proposée par Ville Pulkki (1997) est une répartition géométrique à puissance constante, qui vise à reproduire le front d?onde acoustique, et ce, en projetant le vecteur position de la source sur les vecteurs correspondant aux positions des haut-parleurs. Elle nécessite seulement 2 haut-parleurs actifs par source lorsque l?on se place dans le domaine bidimensionnel et 3 pour un son tridimensionnel.

Bien sûr, il est possible d?utiliser plus de haut-parleurs, comme dans les systèmes Dolby 5.1, 7.1 voire 22.2, le premier nombre correspondant au nombre de haut-parleurs et le second au nombre caissons de basses additionnels. Par exemple, dans le cas du 5.1, on trouve : une stéréophonie avant (+/- 30 degrés), 2 haut-parleurs à l?arrière (+/- 110 degrés), un autre haut-parleur à l?avant (0 degrés) ? utile pour le cinéma où les voix sont souvent positionnées de manière centrée ? et un caisson de basses. Ce sont toutefois des systèmes empiriques, pour des domaines d?application spécifiques (ici le cinéma). Il existe cependant des méthodes plus génériques.

2.4. Ambisonie

L?ambisonie, proposée par Michael Gerzon (1973) et généralisée aux ordres supérieurs (*Higher Order Ambisonics* ou HOA en anglais) par Jérôme Daniel (2001), consiste à reproduire le champ acoustique au niveau de l?auditeur, en décomposant l?espace sur la base des harmoniques sphériques (voir Figure 14).

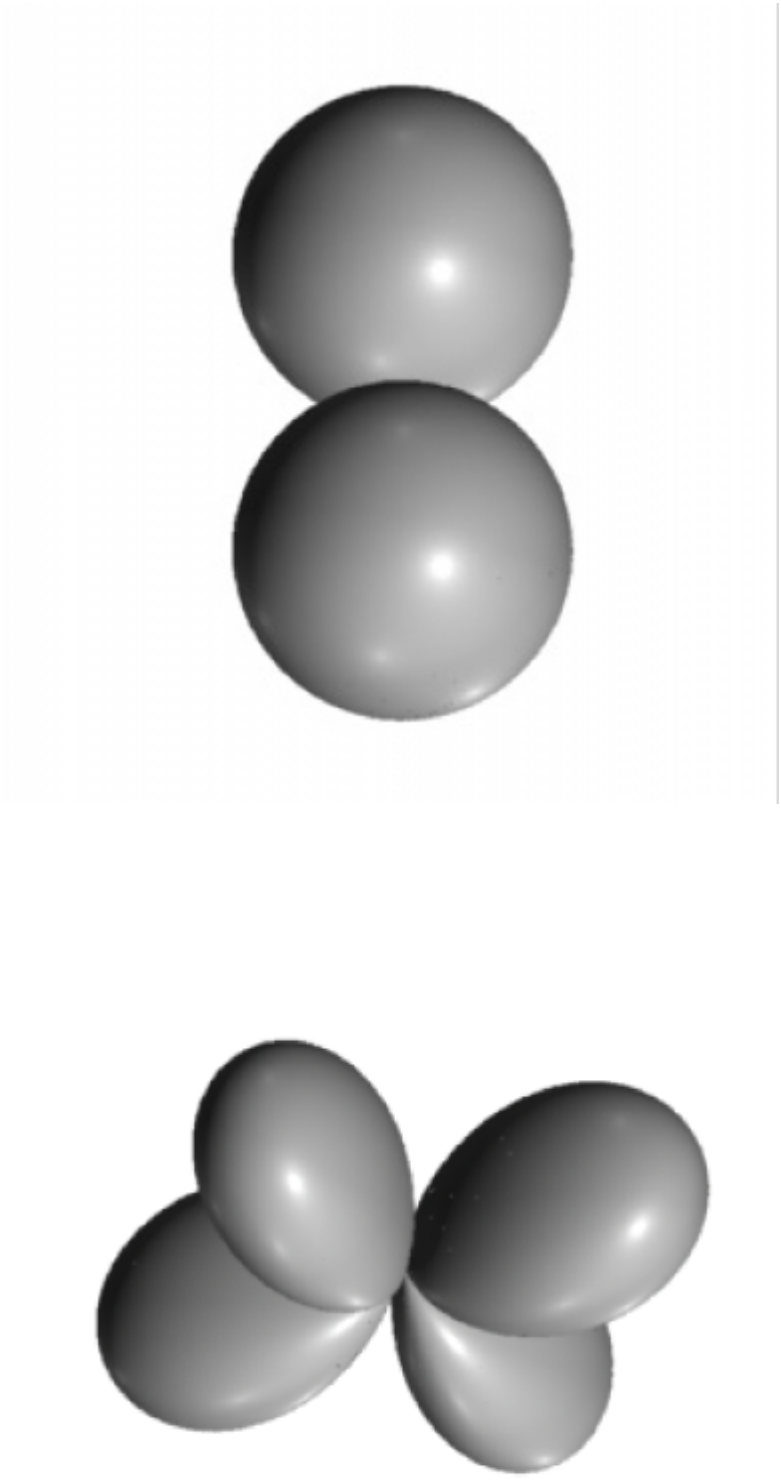
Figure 14. La base des harmoniques sphériques (ici à l?ordre 3)



Source : Franz Zotter et Matthias Frank, « Ambisonic amplitude panning and decoding in higher orders », in *Ambisonics. A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, Cham, Springer, coll. « Springer Topics in Signal Processing », 2019. DOI : https://doi.org/10.1007/978-3-030-17207-7

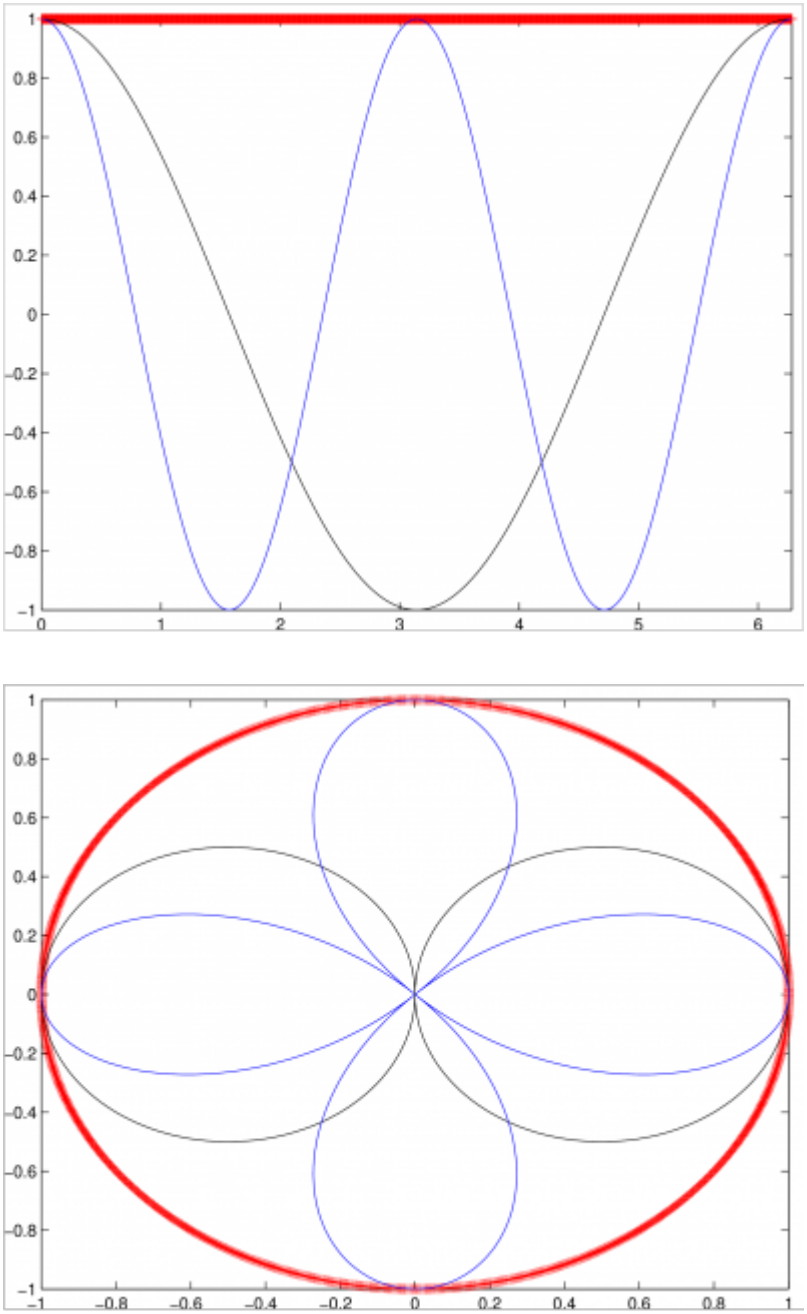
En 3D, ces harmoniques sphériques correspondent à des fonctions de Fourier-Bessel et en 2D (à élévation nulle), il s?agit de la base de Fourier (voir les Figures 15 et 16).

Figure 15. Harmoniques sphériques numéro 0, 1, et 4 (cas 3D)



Source : Sylvain Marchand, 2019

Figure 16. Harmoniques de Fourier



Note : Numéro 0 (en rouge), 1 (en noir) et 2 (en bleu) en représentations cartésienne (en haut) et polaire (en bas)

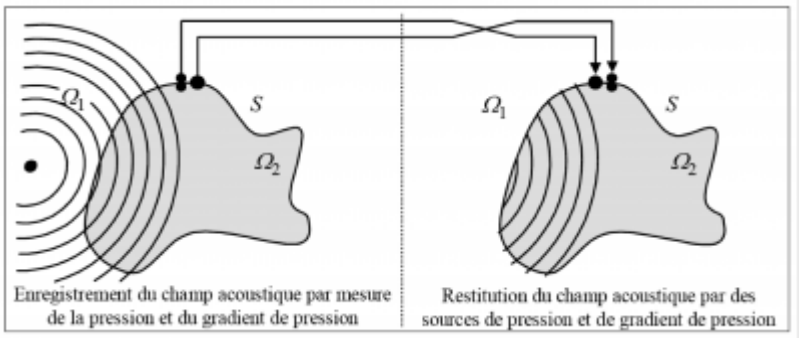
Source : Sylvain Marchand, 2019

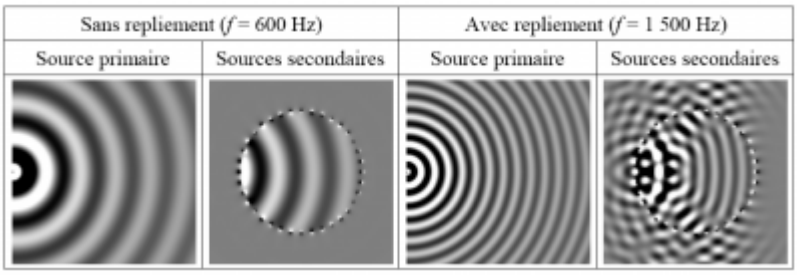
La technique ambisonique repose sur un mécanisme d?encodage/décodage, la captation étant possible (avec des microphones spécifiques comme le SoundField), ainsi que la simulation. De manière simplifiée, il s?agit d?encoder une onde plane pour la source primaire et de décoder des ondes planes pour les sources secondaires (les haut-parleurs), *via* la base des harmoniques sphériques, et ce, à un certain ordre (la base théorique étant de dimension infinie). Notons au passage que l?ordre 0 correspond au cas monophonique. Le problème est le suivant : en théorie, le nombre (et même la configuration) des haut-parleurs est imposé par cet ordre. De nombreux ajustements pratiques (ou ?optimisations?) ont alors été proposés pour lever cette contrainte.

2.5. Holophonie

L?holophonie (*Wave Field Synthesis* ou WFS en anglais) est l?analogue de l?holographie pour les ondes acoustiques. Cette méthode repose sur le principe de Christiaan Huygens (1690) : 1 source primaire est remplacée par *n* sources secondaires. Ce principe a été quantifié par Gustav Kirchhoff et Hermann von Helmholtz (xix^{e} siècle), puis appliqué au son par Augustinus Berkhout (1988). En principe, la méthode reproduit l?ensemble du champ acoustique à l?intérieur d?une zone donnée (voir Figure 17) et l?enregistrement est possible (même s?il faudrait beaucoup de microphones). Il faut néanmoins que les sources secondaires soient suffisamment proches pour pouvoir reconstruire une fréquence *f* (distance inférieure à $c/(2f)$ où *c* est la vitesse de propagation du son). C?est un problème, car cela conduit en pratique à une limitation à $f < 1000$ Hz. De plus, les calculs en machine sont très lourds.

Figure 17. Enregistrement et restitution pour la technique holophonique





Note : En théorie (en haut) et en pratique (en bas), avec phénomène de repliement spatial si les sources secondaires sont trop espacées

Source : Figures 3.5 et 3.6 extraites de Larcher *et al.*, « Techniques de spatialisation des sons », in *Informatique musicale : du signal au signe musical (Traité IC2, série Informatique et systèmes d?information)*, F. Pachet, J.-P. Briot, coord. © Lavoisier, 2004, p. 137 et 138.

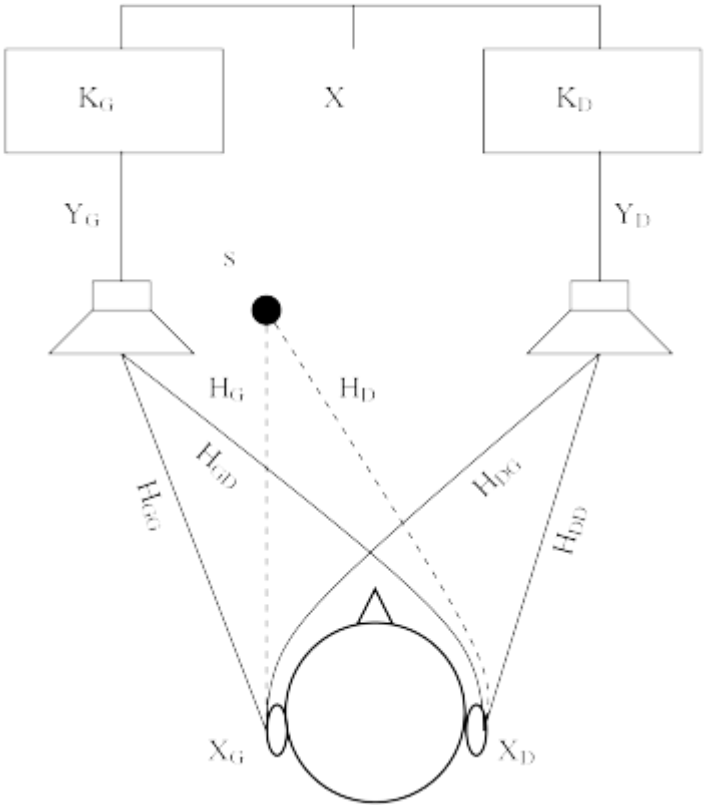
2.6. Vers une approche perceptive

Les approches précédentes sont motivées par la physique, visant la reproduction de l?onde sonore (VBAP) ou bien du champ acoustique au niveau de l?auditeur (ambisonie) ou partout dans l?espace (holophonie).

Ce qui compte en revanche pour un auditeur humain, c?est la perception qu?il a de l?espace. Récemment, nous avons proposé une approche perceptive avec une méthode qui s?intéresse aux chemins acoustiques (obtenus de manière synthétique par les équations de la fin de la section 2) pour reproduire, par la technique transaurale, les indices acoustiques interauraux utilisés par le système auditif pour localiser les sources sonores (voir Figure 18).

Cette méthode, nommée Star (pour *Synthetic Transaural Audio Rendering* ; Mouba et Marchand, 2006 ; Mouba *et. al.*, 2008 ; Méaux et Marchand, 2019), est efficace puisqu?elle ne nécessite qu?une transformée de Fourier rapide par source et une transformée de Fourier inverse par haut-parleur.

Figure 18. Source virtuelle reproduite à l?aide de la technique Star



Source : Sylvain Marchand, 2019

Les coefficients à employer pour les haut-parleurs, K_G et K_D, sont les 2 inconnues du système à 2 équations suivant :

$$X_G = H_G X = H_{GG} K_G X + H_{GD} K_D X$$
$$X_D = H_D X = H_{DG} K_G X + H_{DD} K_D X$$

qui admet donc une solution (le déterminant n?étant pas nul si les haut-parleurs ne sont pas alignés avec la tête de l?auditeur). Des tests subjectifs et objectifs sont en cours pour valider cette approche.

Conclusion

Dans ce qui précède, nous avons évoqué les techniques suivantes :

- la stéréophonie : l?écoute est localisée et l?enregistrement est possible ;
- le binaural : l?écoute se fait avec un casque, l?enregistrement est possible mais il est nécessaire d?individualiser les HRTF ;
- le transaural : l?écoute est localisée et l?enregistrement est possible, mais le calibrage est nécessaire ;
- l?holophonie : l?écoute est libre (mais pas trop près des haut-parleurs) et l?enregistrement est possible (du moins en théorie) ;

- l?ambisonie : l?écoute se fait au niveau du *sweet spot*, l?enregistrement est possible (avec des microphones spécifiques) et il est nécessaire d?augmenter l?ordre (HOA) pour une meilleure résolution spatiale.

À cela, il faut ajouter les lois de *panning*, la technique VBAP et la nouvelle approche Star.

La plupart des techniques de spatialisation du son ont pour objectif de reconstruire l?onde acoustique (partout dans l?espace ? holophonie ?, ou au niveau de l?auditeur ? ambisonie). Certaines concernent l?image spatiale restituée, d?autres modélisent des objets sonores spatiaux (sources ponctuelles et signaux associés, qui se généralisent dans le monde de la production cinématographique). Ces différentes techniques pour le ?son 3D? sont maintenant en phase de normalisation industrielle.

Il reste cependant un espace pour une approche psycho-physique, puisque ce qui compte, en définitive, c?est la sensation de l?auditeur (d?où la nécessité de prendre en considération les mécanismes de la localisation du son). C?est l?occasion d?une recherche interdisciplinaire, entre physique et psychologie, impliquant arts et sciences, puisque l?objectif est la spatialisation de la musique.

Pour citer ce document:
Sylvain Marchand, Â« Une approche perceptive pour la spatialisation du son Â», *RFIM* [En ligne], Numéros, n° 7-8 - Culture du code, Mis Ã jour le 21/12/2020
URL: http://revues.mshparisnord.org/rfim/index.php?id=606