



Algorithme d'upmix ne modifiant pas le timbre

Julien Bréval, Benoît Courribet

Table des matières

1	Introduction	3
2	Principe du <i>M/S processing</i>	4
3	Définition de l'algorithme d'upmix	5
3.1	Conditions d'utilisation.....	5
3.2	Description de l'algorithme.....	5
3.3	Application pratique avec SoX	7
3.4	Downmixabilité.....	8
4	Traitements additionnels	9

1 Introduction

Dans le contexte de l'audiovisuel, il est de plus en plus fréquent d'utiliser une bande son multicanal (5.1, etc.). Or de nombreuses bandes son n'existent qu'en 2 canaux.

Pour des raisons commerciales essentiellement, on veut souvent obtenir une version en 6 canaux d'une bande son à l'origine en 2 canaux.

Bien qu'un remixage manuel à partir des enregistrements audio originaux soit recommandé, cela n'est parfois pas possible en termes de temps de travail ou de coût. Dans ce cas, on a recours à un processus appelé "upmix".

L'upmix est un processus automatique ou semi-automatique (c'est-à-dire paramétrable) permettant d'augmenter le nombre de canaux d'un flux audio.

Dans ce texte, nous présentons un algorithme d'upmix de 2 canaux vers 6 canaux ayant pour principale caractéristique de ne pas modifier le timbre de la bande son d'origine.

2 Principe du *M/S processing*

Le *M/S processing* est une technique de traitement du signal fonctionnant avec des flux audio de 2 canaux différents. L'algorithme d'upmix décrit dans ce texte est basé sur le *M/S processing*, ainsi il est important de rappeler le fonctionnement de cette technique.

Si on appelle X_L et X_R les canaux sur lesquels on veut appliquer un traitement, et Y_L et Y_R les canaux obtenus à la suite du traitement, le principe est le suivant :

- Encodage en M/S ; on introduit un canal $X_M = X_L + X_R$, qui représente la somme des deux canaux d'entrée (ou bien l'information monophonique), ainsi qu'un canal $X_S = X_L - X_R$, qui représente la différence entre les deux canaux (ou bien l'information spatiale)
- Traitement ; on peut appliquer des traitements audio différents sur les canaux X_M et X_S ; on obtient deux canaux appelés ici Y_M et Y_S
- Décodage ; $Y_L = Y_M + Y_S$ et $Y_R = Y_M - Y_S$.

3 Définition de l'algorithme d'upmix

3.1 Conditions d'utilisation

Cet algorithme permet d'obtenir un upmix d'un flux audio de 2 canaux vers 6 canaux ("5.1 surround"). Cependant, deux conditions sont nécessaires pour son bon fonctionnement :

- les deux canaux d'entrée doivent être différents
- le downmix en mono des 2 canaux d'entrée doit être acceptable.

3.2 Description de l'algorithme

On appelle X_L et X_R les canaux d'entrée. Les canaux de sortie sont appelés Y_L , Y_R , Y_C , Y_{LFE} , Y_{SL} et Y_{SR} .

Les valeurs a , b , c et d sont des coefficients de mixage positifs, dont les valeurs ont été trouvées empiriquement en effectuant des tests sur des sources sonores très diverses.

Soient les canaux $M = X_L + X_R$ et $S = X_L - X_R$.

On obtient les canaux Y_L et Y_R en appliquant des coefficients de mixage différents sur M et S :

- $Y_L = aM + bS$
- $Y_R = aM - bS$

Si on choisit les valeurs $a = 0.295$ et $b = 0.405$, comme $b > a$, l'espace sonore obtenu par la diffusion de Y_L et Y_R est plus large que celui produit par X_L et X_R , ce qui permet d'ajouter un canal Y_C sans que l'image sonore frontale de la version 6 canaux ne soit plus étroite que celle de la version deux canaux :

- $Y_C = cM$

Le coefficient de mixage c permet de ne pas obtenir un volume sonore trop important pour le canal central (ce qui réduirait également la largeur stéréophonique avant). Nous avons choisi $c = 0.354$.

Le canal LFE est destiné à la diffusion ou au renfort d'effets spéciaux sonores ayant besoin de beaucoup de puissance dans les graves (explosions, foudre, tremblements de terre, catastrophes, désordre, etc.). Ce canal est diffusé uniquement par un ou plusieurs subwoofers. Cependant, dans certains systèmes 5.1, les enceintes L, R, C, SL et SR ne sont pas large bande et ne peuvent notamment pas reproduire les graves. Dans ce cas, on utilise donc le subwoofer pour reproduire également les graves des canaux L, R, C, SL et SR, grâce à la technique du *bass management*. Dans certains systèmes 5.1 domestiques, il n'y a même pas de canal LFE et le subwoofer est uniquement utilisé pour le *bass management*.

Ainsi, nous avons choisi de créer un canal LFE diffusant les basses des canaux L et R en utilisant la technique du *bass management* :

- $Y_{LFE} = LPF(dM)$, avec LPF désignant un filtre passe-bas d'ordre 2, de fréquence de coupure égale à 100Hz et de facteur de qualité égal à 0.71 ; on choisit $d = 0.5$

Si le dispositif de diffusion possède son propre système de *bass management*, il ne faut pas utiliser le canal LFE ainsi obtenu. Une variante intéressante consisterait à utiliser un filtre passe-bas ayant une fréquence de coupure beaucoup plus grave (de l'ordre de 30Hz à 50Hz). Dans ce cas, seules les basses les plus graves de X_L et X_R seraient diffusées dans Y_{LFE} .

Etant donnée la diversité des systèmes 5.1 existants, il est difficile de proposer une solution satisfaisante dans tous les cas pour Y_{LFE} .

Les canaux Y_{SL} et Y_{SR} sont obtenus de manière analogue aux canaux Y_L et Y_R , mais en utilisant des coefficients de mixage différents :

- $Y_{SL} = eM + fS$
- $Y_{SR} = eM - fS$

Si on choisit les valeurs $e = 0.225$ et $f = 0.445$, comme $e < a$ et $f > b$, l'espace sonore obtenu par la diffusion de Y_{SL} et Y_{SR} est encore plus large que celui produit par Y_L et Y_R , ce qui est cohérent par rapport à la position habituelle des enceintes de surround.

3.3 Application pratique avec SoX

SoX¹ est un logiciel en ligne de commande multiplateformes (Windows, Linux, Mac OS X), libre et gratuit, qui peut convertir des fichiers audionumériques d'un codec ou d'un format à un autre. Par ailleurs, il permet d'appliquer divers traitements sur ces fichiers.

On appelle "source.wav" le fichier audio à partir duquel on veut réaliser un upmix. Il s'agit ici obligatoirement d'un fichier audio de deux canaux entrelacés. L'upmix correspond à 6 fichiers monophoniques appelés "upmix-L.wav", "upmix-R.wav", "upmix-C.wav", "upmix-LFE.wav", "upmix-SL.wav" et "upmix-SR.wav".

Voici le calcul de l'upmix :

```
sox source.wav upmix-L.wav remix 1v0.7,2v-0.11
sox source.wav upmix-R.wav remix 1v-0.11,2v0.7
sox source.wav upmix-C.wav remix 1v0.354,2v0.354
sox source.wav upmix-LFE.wav remix 1v0.5,2v0.5 lowpass -2 100 0.71
sox source.wav upmix-LS.wav remix 1v0.67,2v-0.22
sox source.wav upmix-RS.wav remix 1v-0.22,2v0.67
```

Le script *hsox*² qui sert de moteur de traitements audionumériques dans l'éditeur audio de HD3D utilise cette technique pour calculer les upmixes.

¹ <http://sox.sourceforge.net/>

² Voir le document *Editeur audio – Documentation de hsox* (également publié dans le cadre du projet HD3D)

3.4 Downmixabilité

Dans le contexte actuel, il est recommandé de produire des bandes son dont le downmix de 6 canaux vers 2 canaux (voire vers 1 canal) est acceptable³.

On montre mathématiquement que si le downmix en mono des canaux \mathbf{X}_L et \mathbf{X}_R est acceptable, alors l'algorithme d'upmix décrit ci-dessus produit un flux de 6 canaux dont le downmix vers deux canaux ou vers 1 canal est acceptable.

En effet, si on appelle \mathbf{Z}_L et \mathbf{Z}_R les canaux obtenus au downmix vers 2 canaux et g , h les coefficients de downmix (positifs) appliqués aux canaux C et de surround, on obtient les relations suivantes :

- $\mathbf{Z}_L = \mathbf{Y}_L + g\mathbf{Y}_C + h\mathbf{Y}_{SL}$
- $\mathbf{Z}_R = \mathbf{Y}_R + g\mathbf{Y}_C + h\mathbf{Y}_{SR}$

Ce qui est équivalent à :

- $\mathbf{Z}_L = (a+b+gc+he+hf)\mathbf{X}_L + (a-b+gc+he-hf)\mathbf{X}_R$
- $\mathbf{Z}_R = (a-b+gc+he-hf)\mathbf{X}_L + (a+b+gc+he+hf)\mathbf{X}_R$

Si on utilise des coefficients de downmix g et h égaux à 1, on obtient :

- $\mathbf{Z}_L = 1.724\mathbf{X}_L + 0.024\mathbf{X}_R$
- $\mathbf{Z}_R = 0.024\mathbf{X}_L + 1.724\mathbf{X}_R$

Ceci montre que le downmix vers 2 canaux obtenu à partir de l'upmix possède une spatialisation légèrement plus étroite que le fichier de 2 canaux d'origine mais reste très acceptable.

Idéalement, il faudrait obtenir $\mathbf{Z}_L = \mathbf{X}_L$ et $\mathbf{Z}_R = \mathbf{X}_R$. Cependant, pour cela, il faudrait que $a-b+gc+h(e-f) = 0$. Or on ne peut pas satisfaire cette condition dans le cas général car g et h ne sont pas connus a priori.

Quand au downmix en mono de l'upmix, on obtient :

- $\mathbf{Z}_M = (2a + 2gc + 2he)\mathbf{X}_L + (2a + 2gc + 2he)\mathbf{X}_R = (2a + 2gc + 2he)(\mathbf{X}_L + \mathbf{X}_R)$

Ainsi, le downmix en mono est parfait à un coefficient de mixage positif près.

³ Voir le document *Downmix – tests audio* (également publié dans le cadre du projet HD3D)

4 Traitements additionnels

Les coefficients de mixage a , b , c , d , e et f ont été choisis empiriquement de manière à convenir à un grand nombre de fichiers source. Cependant, il est possible que l'équilibre de l'upmix ne soit pas tout à fait satisfaisant dans certains cas. Ainsi, on peut recommander de modifier les volumes des 6 pistes obtenues à la fin de l'upmix.

Par ailleurs, dans le cas d'une bande son contenant une source sonore principale qu'il faut mettre en avant (voix des acteurs, instrument soliste, etc.), on peut filtrer fréquemment le canal central de manière à favoriser cette source par rapport aux autres.

De même, les canaux de surround peuvent être filtrés fréquemment afin de modifier la spatialisation obtenue. On peut même ajouter un léger retard (environ 10 millisecondes) à ces canaux afin d'obtenir un rendu spatial plus immersif, mais dans ce cas il faut s'assurer que le downmix vers 2 canaux (voire en mono) est acceptable.