Lidwine **Hô**
France télévisions - innovations & développements

Hervé **Dejardin**
Radio France Innovation

TUTORIEL

Production audio pour diffusion Youtube 360°

Novembre 2016

Ce tutoriel a pour objectif de vous aider à produire des fichiers en audio spatialisé qui seront ensuite assemblés avec des fichiers vidéo afin d'être diffusés sur YouTube 360°.
Dans ce document, nous partageons la méthode et la liste des outils que nous avons utilisées.
Cette méthode nous a semblée la plus simple à mettre en œuvre.
Nous avons utilisé Windows7 sur un PC et les explications qui suivent se limitent à ce système d'exploitation. Il est possible de reproduire cette méthode avec un Mac sous OS X. Mais nous n'en décrivons ni les étapes, ni les procédures spécifiques à cet OS.
Ce tutoriel n'explique pas comment produire des images à 360°.
Ce tutoriel n'est donc en aucun cas exhaustif.

SOMMAIRE

Principe de la chaîne de production et de diffusion

La chaine de diffusion YouTube 360° prend en charge le téléchargement et la lecture de vidéos sphériques à 360°.
Vidéo sphérique à 360° sous-entend une image et un son à 360° interactifs.

Nous avons testé ces vidéos à 360° sur des ordinateurs (Win et OSX) avec le navigateur Chrome ainsi que sur Smartphone et tablette sous Android avec l'application YouTube.
Avec l'application YouTube pour Android vous pouvez regarder des vidéos à 360° avec un casque de réalité virtuelle adapté à votre Smartphone ou encore avec un Cardboard.

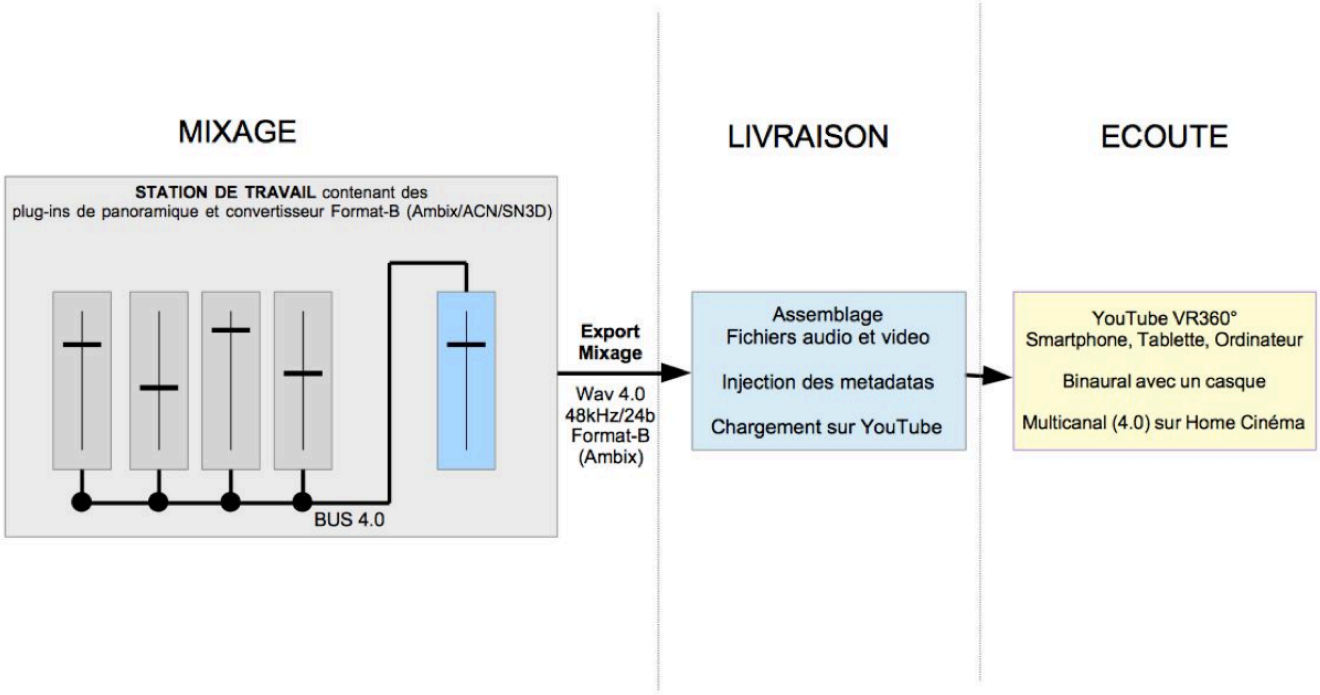
L'audio est prévue pour une écoute immersive en binaural avec un casque. Cette écoute peut être interactive et ainsi permettre de simuler les mouvements de rotation de la tête en azimut (mouvement de rotation gauche et droit) et en élévation (mouvement de rotation haut et bas).

Il est également possible d'écouter l'audio spatialisé des contenus YouTube360° sur un home cinéma en 5.1. (Le format audio utilisé ne permet pas une grande séparation des canaux. La qualité en 5.1 ne sera donc pas très bonne).
Pour une bonne qualité d'expérience immersive, nous recommandons l'écoute binaurale avec des écouteurs (casque audio) et un casque de réalité virtuelle .

Comme décrit sur le synoptique suivant, le signal audio est dans un format appelé ambisonique. Ce format permet une distribution de l'audio spatialisé avec uniquement quatre canaux. Il permet également une économie des ressources processeur nécessaires à la simulation des mouvements de rotation de la tête.

PS: A ce jour, seule l'application YouTube sous Android permet d'écouter en binaural.

Chaîne de production YouTube 360



Quelques explications sur la reproduction audio binaural et sur le format ambisonique utilisés par YouTube

Binaural Audio

La diffusion en audio binaural fonctionne à l'aide de fonctions HRTFs (Head Related Transfer Functions). Les HRTFs filtrent le signal pour recréer les repères complexes qui nous aident à localiser les sons.

Notre cerveau utilise essentiellement trois indices pour se repérer dans l'espace :

- Le premier indice est la différence d'intensité entre nos deux oreilles. L'oreille qui perçoit le plus d'intensité étant l'oreille la plus proche de la source audio.

- Le second indice est la différence de temps que met le son pour arriver à nos deux oreilles. L'oreille la plus proche de la source sonore perçoit celle-ci en premier. Si le son arrive avec la même intensité et au même instant dans nos deux oreilles, alors la source sonore est soit devant nous à 0° soit derrière nous à 180°.

Nous utilisons ces deux premiers indices pour localiser la source sur le plan horizontal.

- Le troisième indice est le filtrage de la source audio engendré par notre morphologie.

En effet, le spectre de la source sonore est modifié par l'incidence de nos épaules, de notre tête, de nos pavillons d'oreilles...
Nous utilisons ce troisième indice pour localiser la source sur le plan médian (avant arrière, haut bas).

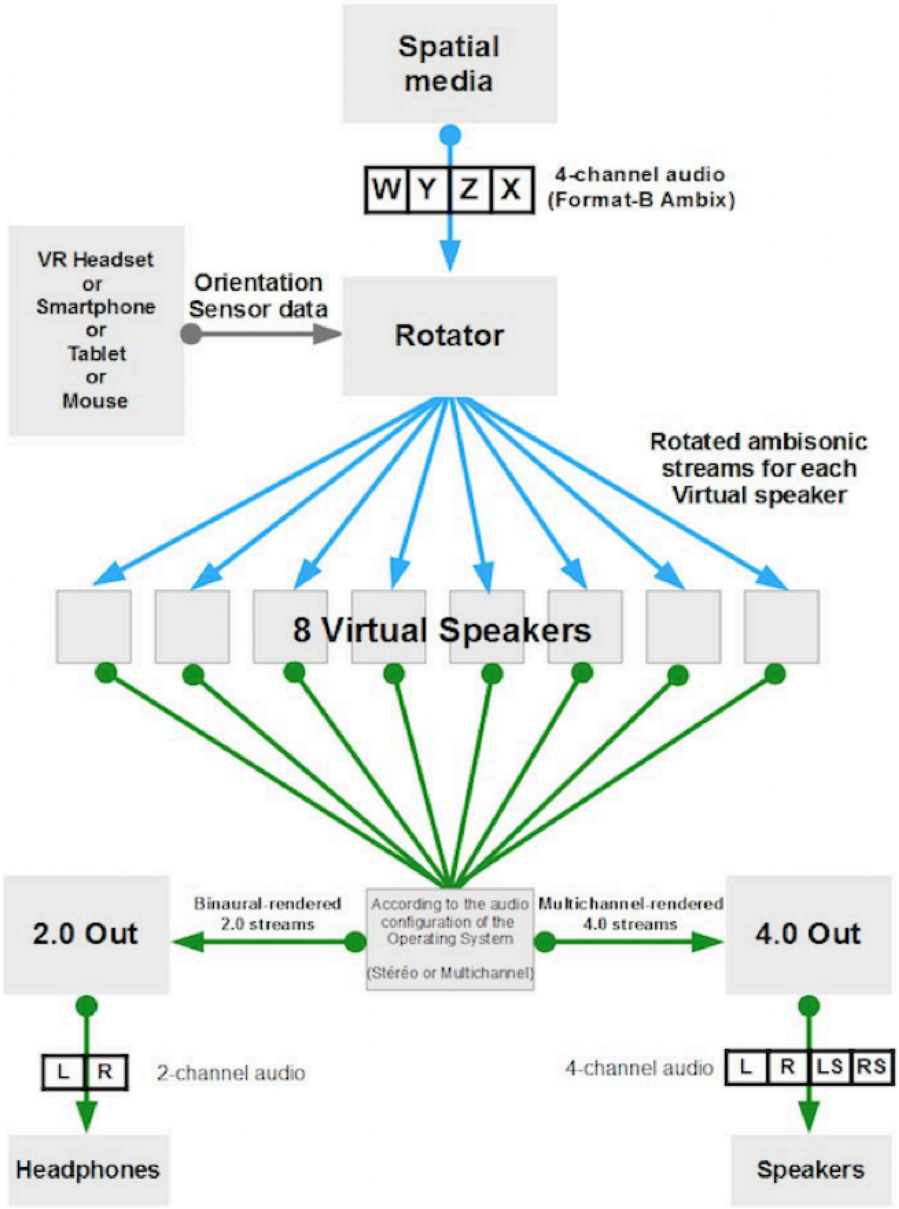
Afin de lever les ambiguïtés de localisation, nous tournons la tête et nous essayons constamment d'amener les sons (ou les objets qui les produisent) dans notre champ de vision.
La qualité d'expérience d'écoute de l'audio binaurale peut-être grandement améliorée avec des capteurs de suivi de mouvement de tête. Ce sont ces capteurs qui sont utilisés dans la réalité virtuelle.

Ainsi, pour l'écoute avec un casque, nous pouvons reproduire une scène sonore en « audio 3D » qui apparaît devant l'auditeur en fonction de l'orientation de sa tête.

Ambisonique et format-B

On peut faire une analogie entre le format ambisonique et la vidéo à 360° : On enregistre toute la scène, mais on ne regardera que dans une direction à la fois : on choisit son point de vue et d'écoute en temps réel.
Le format ambisonique apparaît aujourd'hui comme un format incontournable.
C'est un format de « description » d'une scène audio 3D, qui peut se décoder dans n'importe quel format de restitution, il est très flexible.
L'ambisonique d'ordre 1 dit Format B ou FOA (First Order Ambisonic) se compose de 4 canaux audio. On parle de HOA (High Order Ambisonic) pour les formats ambisoniques d'ordre plus élevé (9 canaux pour l'ordre 2, 16 pour l'ordre 3...)
La formule qui permet de calculer le nombre de canaux en fonction de l'ordre est pour un signal audio 3D $(ordre+1)^2$ = Nombre de canaux.
Plus l'ordre est élevé plus la « description spatiale » est précise.

Ces canaux qui composent le format ambisonique sont des canaux audio destinés à être décodés par un calcul d'harmoniques sphériques.
Il existe deux principaux standards de Format B : le Fuma et l'Ambix. L'Ambix est actuellement le plus courant (en ce qui concerne les formats d'export) notamment parce que celui-ci est utilisé par les Players 360° de YouTube et FaceBook.
Il est assez simple de passer du format Ambix au FuMa et vice versa, grâce à des plugins (voir liens en fin de document)
Le format ambisonique ne se préoccupe pas de savoir de quoi est constituée la scène, c'est un format « scene-based » (basé sur la description de la scène audio 3D).
Les différents avantages du format B (Ambisonic ordre 1 ou FOA) sont qu'il est peu gourmand en nombre de pistes et les calculs sont assez simples : on peut faire des rotations de la scène audio très facilement.
Le format ambisonique n'est pas qu'un format pivot ou de restitution, il permet aussi d'enregistrer des scènes grâce à des micros dits ambisoniques. A l'ordre 1, les différents canaux audio obtenus sont au format A : format non standardisé. Il faut donc les convertir du format A au format B grâce à un outil en général fourni par le constructeur du micro. Cet outil matrice et traite les signaux produits par les quatre capsules microphoniques à directivité cardioïde.
Le format B permet de choisir la direction d'écoute mais aussi le format d'écoute (mono, stéréo, 4.0, 5.1, binaural...)
Il suffit que le « player » sache décoder les informations contenues dans le format B pour le convertir au format de restitution souhaité.

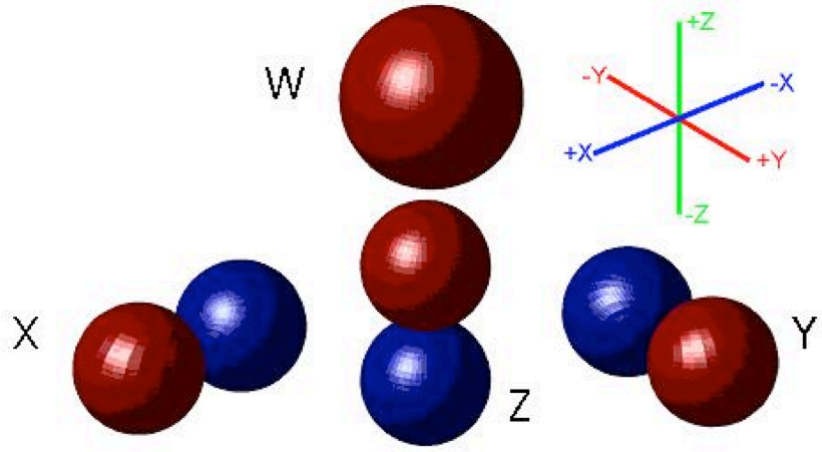


Ci-contre un synoptique de traitement du signal ambisonique permettant les mouvements de rotation de la tête avec une sortie audio au format binaural ou multicanal. Cette chaine de traitement est similaire à celle utilisée par YouTube360°.

Beaucoup d'outils de manipulation audio à 360° proposent le format ambisonique comme format pivot de leur moteur de rendu.
La précision étant meilleure avec des ordres plus élevés, les chaines de traitement sont parfois à l'ordre 2, 3, 4, ...7,...
Sur YouTube360°, c'est le format-B (ordre1) qui est exclusivement utilisé.

Caractéristiques du format B

Le Format-B est donc archivé sur quatre canaux audio. Ces quatre canaux sont nommés W,X,Y,Z dont voici la description.
Pour un auditeur au centre du champ sonore, le champ ambisonique doit être défini sur un ensemble d'axes 3D (X, Y, Z).
Par convention, X représente l'axe d'avant en arrière, Y de gauche à droite et Z de haut en bas.
Les canaux X, Y, Z contiennent des informations sur le champ sonore le long de l'axe correspondant (ils représentent ce qu'un microphone bidirectionnel peut capter le long de cet axe).
Le canal W peut être considéré comme un canal omni directionnel.



Les deux principaux standards du format B

1. Le format B connu sous le nom « FuMa » utilise un ordre des canaux et une pondération qui ont été proposés par deux chercheurs (Furse et Malham). Nous appellerons ce standard Format-B FuMa. Le format B (FuMa) répertorie les canaux dans l'ordre: W, X, Y, Z. Le canal W est atténué de 3dB.

2. Le format B connu sous le nom « AmbiX » est caractérisé par un ordre et une pondération des canaux différents de la convention « FuMa ». Nous appellerons ce standard Format-B AmbiX. Le format B (AmbiX) répertorie les canaux en utilisant la numérotation des canaux ambisoniques (ACN) : W, Y, Z, X. Les canaux utilisent la normalisation SN3D (Pour l'ambisonique d'ordre 1 cela signifie simplement que les quatre canaux ont une normalisation de gain uniforme).

YouTube utilise le format-B AmbiX.

1 Enregistrement

Comme vu précédemment, il est possible d'enregistrer directement en ambisonique au format B.
Des microphones ainsi que des enregistreurs prévus pour le format B sont disponibles à des prix raisonnables.
Il faudra s'assurer que le format B est bien au standard AmbiX.
Des plug-in de conversion du standard FuMa vers AmbiX existent (voir fin du document).

Les enregistrements au format audio orienté canal (mono, stéréo, quadriphonique, 5.1...) peuvent être convertis au format ambisonique.
Cette conversion peut se faire avec des plug-in (format VST, AAX, AU...) installés sur un logiciel de mixage et d'édition audio (Cubase, Pyramix, Reaper, Logic, Pro-tools...).

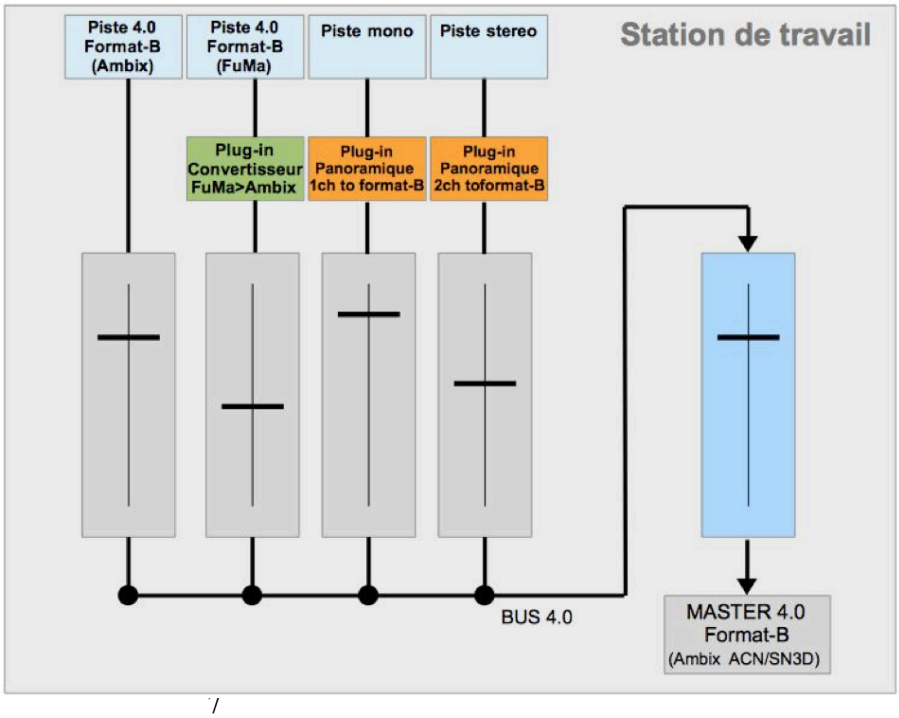
2 Mixage

Le mixage peut-être réalisé avec des logiciels de mixage et d'édition audio permettant de mixer sur des bus quatre canaux.

Il faudra créer un projet avec un bus principal (master) qui devra être au format multicanal 4.0.

Selon le format de la piste audio alimentant le bus principal, il faudra insérer sur la chaine de traitement des plug-in permettant de convertir ou de traiter le signal au format B (Ambix)

Voici un exemple simple de mixage ambisonique réalisé à partir de différents formats de sources audio.
Sur le synoptique de mixage suivant, sont mélangés une piste au format B (Ambix), une piste au format B (FuMa), une piste monophonique et une piste stéréophonique.



La sortie principale (master) sera donc en 4.0 au format B. C'est cette sortie qu'il faudra utiliser pour l'export du mixage.

3 Export audio

Une fois le mixage terminé, il faut exporter l'audio spatialisé au format B (AmbiX).

Pour une qualité optimale, nous recommandons de réaliser l'export au format audio suivant :
Format Wav 48 kHz / 24 bits multicanal 4.0.

Niveaux audio pour l'export

Il n'est actuellement pas très aisé de prédire le niveau d'écoute sur le player de YouTube 360°.
Car le niveau mesuré sur le master 4.0 (format B AmbiX) ne sera pas celui de la sortie audio binaural du player de YouTube360.
Les traitements et les filtres (HRTF) utilisés par le player lors de l'étape d'encodage du format B en binaural modifient le niveau de la modulation en sortie.
Actuellement, nous ne disposons pas des informations nécessaires à la parfaite simulation de la chaine de traitement développée et utilisée par YouTube.

L'unique moyen (à notre connaissance) d'obtenir le niveau de sortie en binaural du player YouTube360 est de simuler celui-ci avec le plug-in AmbiHead développé par NoiseMakers (voir liens).
Ce plug-in reproduit de manière assez similaire l'encodage en binaural du signal ambisonique réalisé par le player YouTube360.
Les quelques mesures que nous avons réalisées montrent que le niveau de sortie binaural du plug-in est assez proche du niveau de sortie binaural du player YouTube.
NB : Nous précisons qu'il s'agit à ce jour d'une approximation.

Nous recommandons un niveau de sonie de -16LUFS avec un niveau de crêtes réelles à -3dBTP pour la sortie audio en binaural.

4 Assemblage audio vidéo

Il s'agit d'assembler le fichier wav 48 kHz / 24 bits entrelacé 4.0 (contenant l'audio au Format B) et un fichier vidéo mp4 avec une fréquence d'image de 24, 25, 30, 48, 50 ou 60 images par seconde et un format d'image 16:9.

Les explications qui vont suivre ne sont valables que sous Windows (Win7 ou supérieur).
NB : Ces actions peuvent également être réalisées sous OSX. Nous n'en ferons pas la description.

L'assemblage des fichiers audio et vidéo à destination de YouTube 360° doit être réalisé à l'aide de la librairie ffmpeg.

C'est actuellement, à notre connaissance, la seule solution d'assemblage (gratuite) qui fonctionne parfaitement.
Car avec les autres méthodes que nous avons essayées, lors de l'injection des metadatas (voir rubrique suivante), le logiciel fourni par YouTube « refusait » de valider l'option audio 360.
Le succès de ces nouveaux formats va surement motiver des développeurs à produire des logiciels qui permettront de réaliser cette opération plus simplement.

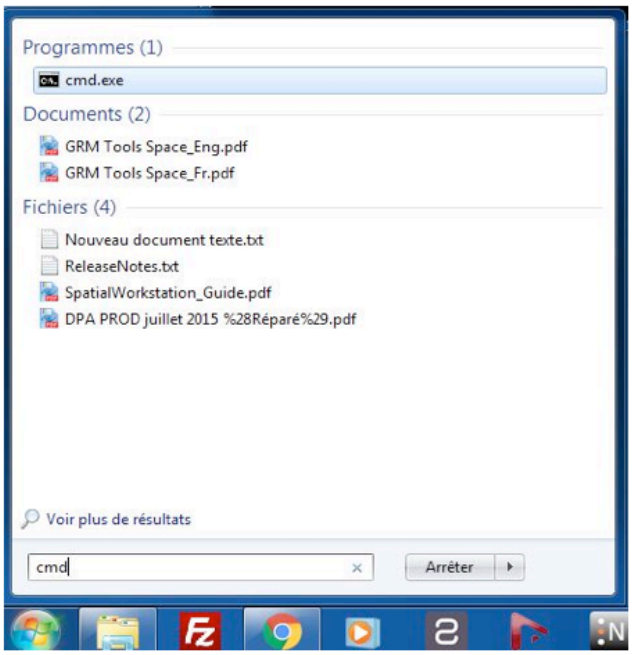
La librairie ffmpeg (Version Windows) peut être téléchargée à l'adresse du lien suivant.

Lien téléchargement ffmpeg
https://ffmpeg.org/download.html

Pour l'installation sous Windows de la librairie ffmpeg, suivre le lien vers le tutoriel suivant :
Lien Tutoriel installation FFMPEG Windows
http://fr.wikihow.com/installer-FFmpeg-sur-Windows

Ce qui suit, sous entend que la librairie ffmpeg est correctement installée sur votre ordinateur utilisant le système d'exploitation Windows 7 ou supérieur.

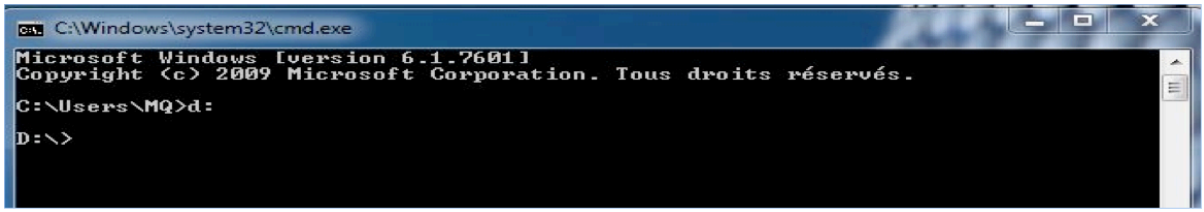
Il vous faut appeler une fenêtre de commande sous Windows.
Dans le menu démarrer, entrez **cmd** dans l'invitation « recherchez les programmes et fichiers ».



C'est par des lignes de commande entrées dans cette fenêtre que vous allez pouvoir utiliser la librairie ffmpeg pour l'assemblage des fichiers vidéo et audio.

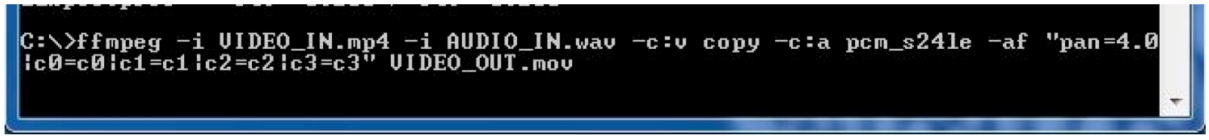
Afin de simplifier les lignes de commande, nous conseillons de réaliser les assemblages de fichiers se trouvant à la racine d'un disque. Plus l'arborescence sera complexe, plus l'adresse du fichier sera longue et fastidieuse à entrer.

Si les fichiers se trouvent à la racine d'un disque, il suffira d'entrer la lettre du disque puis valider par la touche « Entrée » de votre clavier.
Pour des fichiers se trouvant à la racine du disque D, entrer d**:** dans la fenêtre de commande (cmd) puis valider par la touche « Entrée ».



Entrer la commande permettant l'assemblage avec les noms des fichiers vidéo et audio correspondants :

ffmpeg -i VIDEO_IN.mp4 -i AUDIO_IN.wav -c:v copy -c:a pcm_s24le -af "pan=4.0|c0=c0|c1=c1|c2=c2|c3=c3" VIDEO_OUT.mov



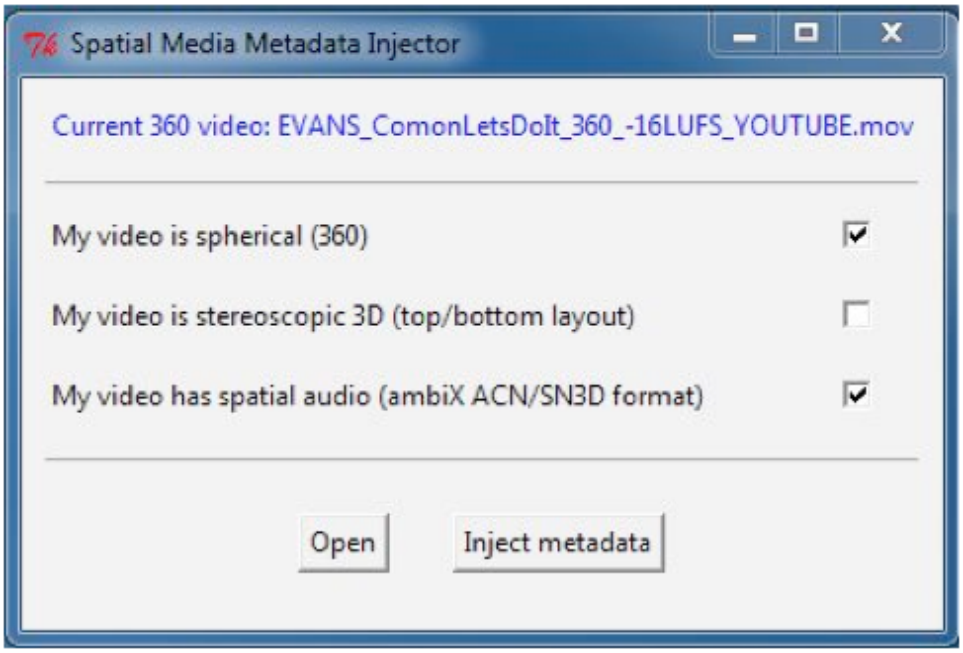
Lancer cette commande en appuyant sur la touche « Entrée »
Le fichier « assemblé » est désormais disponible à l'adresse que vous avez indiquée.

5 Injection des metadatas

Une dernière opération est nécessaire avant l'upload du fichier sur les serveurs de YouTube.
Celle-ci consiste à injecter des metadatas dans le fichier à l'aide d'une application (Spatial Media MetaData injector).
Ces metadatas vont permettre au serveur d'identifier si le fichier contient une image à 360° ou stéréoscopique et si l'audio est spatialisé.

Cette application est gratuite, pour windows elle est disponible à cette adresse
https://github.com/google/spatial-media/releases/download/v2.0/360.Video.Metadata.Tool.win.zip

Une fois téléchargée et dézippée, l'application vous permet de sélectionner le fichier à traiter.
Puis il faut valider les différentes options (Spherical 360, Stereoscopic 3D, Spatial Audio).
Comme sur l'image suivante, dans le cas d'une vidéo 360° sphérique avec de l'audio spatialisé.



Valider en cliquant sur Inject Metadata.
L'application génère une copie du fichier sélectionné avec les metadatas injectées. Elle ajoute le suffixe « injected » au nom du fichier.

6 Livraison

Il ne reste plus qu'à mettre en ligne les fichiers sur un compte YouTube suivant la procédure habituelle.

Ensuite il faudra après l'import du fichier et son traitement, attendre entre 15 et 30 minutes avant que la vidéo soit totalement fonctionnelle à 360°.

7 Écoute

Afin d'être sure de bénéficier du son et de l'image à 360° avec interactivité, nous vous recommandons :

- L'application YouTube installée sur un Smartphone ou une tablette récente dont l'OS est Android.

- Le navigateur Chrome sur Mac et PC. Dans ce cas vous pourrez écouter en stéréo ou en 5 .1 sur hauts parleurs. En 5.1, l'image et le son tournent simultanément sous l'action de la souris.

PS: A ce jour, seule l'application YouTube sous Android permet d'écouter en binaural.

Nous avons expérimenté l'immersion à 360° avec un casque audio ainsi qu'un casque de réalité virtuelle Homido. Nous avions placé dans le casque Homido, un Smartphone Galaxy S5.
Celui-ci utilisait l'application YouTube pour lire le contenu audio et vidéo à 360°.

L'immersion sera totale avec des « lunettes de réalité virtuelle de bonne qualité et sachant se faire oublier » ... ainsi qu'avec un bon casque audio.

Lignes de commandes ffmpeg

Obtenir la version de la librairie ffmpeg installée.
ffmpeg -version

Changement d'encapsulage .mp4 vers .mov
ffmpeg -i input_file.mp4 -acodec copy -vcodec copy -f mov output_file.mov

Encodage H264 (420p) encapsulage .mov
ffmpeg -y -probesize 5000000 -i YOUR_INPUT_FILE -c:v libx264 -profile:v main -vendor ap10 -pix_fmt yuv420p -an YOUR_OUTPUT_FILE.mov

Assemblage fichier video avec fichier audio 4.0 wav
ffmpeg -i video_file_in.mov -i audio_file_in.wav -c:v copy -c:a pcm_s24le -af "pan=4.0|c0=c0|c1=c1|c2=c2|c3=c3" video_audio_file_out.mov

Liens

YouTube Help
https://support.google.com/youtube/topic/2888648?hl=en&ref_topic=16547

(YouTube Help) Déposer une vidéo à 360° sur YouTube
https://support.google.com/youtube/answer/6178631?hl=en&ref_topic=2888648

(YouTube Help) Audio spatialisé pour une video à 360° sur YouTube
https://support.google.com/youtube/answer/6395969?hl=en&ref_topic=2888648

Github Google
https://github.com/google/spatial-media

Format Ambisonic
https://en.wikipedia.org/wiki/Ambisonics

Téléchargement ffmpeg
https://ffmpeg.org/download.html

Installation ffmpeg sous Windows
http://fr.wikihow.com/installer-FFmpeg-sur-Windows

Documentation ffmpeg
http://ffmpeg.org/documentation.html

Spatial Media MetaData Injector
Pour Win
https://github.com/google/spatial-media/releases/download/v2.0/360.Video.Metadata.Tool.win.zip
Pour OSX
https://github.com/google/spatial-media/releases/download/v2.0/360.Video.Metadata.Tool.mac.zip

Spatial WorkStation Facebook
https://facebook360.fb.com/spatial-workstation/

Blog Bruce Wiggins
http://www.brucewiggins.co.uk/

Plug-in ambisonique
Collection ambiX
http://www.matthiaskronlachner.com/?p=2015
NoiseMakers
http://www.noisemakers.fr/