

École nationale supérieure Louis-Lumière
Promotion 2018
Section son

Mémoire de fin d'études

Enregistrement et traitement de métadonnées dynamiques dans le flux de travail audio numérique

Jules Laurin

Sous la direction de

Sylvain Lambinet

Pascal Grillère

Yves-Marie Omnes

Rapporteur

Jean Rouchouse

« Car c'est dans l'octet que toutes nos émotions sonores passent, tout compte fait... »¹

Yves-Marie Omnes

¹ **Éric LONNI**, « Compte-rendu - L'atelier « Que se cache-t-il derrière les fichiers wave et les metadatas » », Association Française du Son à l'Image, septembre 2014.

Remerciements

Je remercie mes directeurs de mémoire, Sylvain Lambinet, Pascal Grillère et Yves-Marie Omnes qui m'ont accompagné et dirigé tout au long de ce travail. Sans leurs conseils et leurs bienveillances, de nombreux points clés de ce mémoire n'auraient pu voir le jour.

Je tiens aussi à remercier les professionnels qui ont accepté de répondre à mon questionnaire dont les entretiens ont réellement fait avancer ma réflexion autour des métiers de la chaîne du son. Merci à Lucien Balibar, Niels Barletta, Katia Boutin, Laure-Anne Darras, Damien Guillaume, François Guillaume, François de Morant et Benjamin Rosier.

Enfin, merci à mes proches, à ma famille et à la promotion 2018 qui m'ont toujours soutenu de près ou de loin et qui m'ont apporté leur aide lors de la rédaction de ce mémoire et dans la réalisation de la partie pratique.

Résumé

Dans le cadre d'un tournage au cinéma, la matière sonore délivrée par le chef opérateur du son se présente sous la forme d'un ensemble de fichiers audio nommé *rushes*. Dans le but de faciliter l'échange de données entre l'étape de tournage et de post-production, les données du son sont référencées et intimement liées à des paramètres supplémentaires, les métadonnées. Ces dernières sont tant informatives que descriptives. Elles renseignent sur différents paramètres essentiels lors de l'exploitation du fichier plusieurs mois après le tournage. Au regard des métadonnées existantes et normalisées, les traitements dynamiques et fréquentiels réalisés en direct au moment de la prise de son ne sont pas retranscrits sous forme de métadonnées. Les intentions du chef opérateur du son sont appliquées directement au signal enregistré de manière définitive et ne sont pas référencées sous forme de métadonnées.

Or, les technologies actuelles permettent de réaliser les traitements de manière non-destructive en employant des courbes d'automatisation. Dans le cadre de ce mémoire, nous proposerons de nouvelles métadonnées dans le but de répondre à différentes problématiques liées à la non-exploitation de données essentielles à la fabrication de la bande son d'un film. Le but de ces recherches est de présenter les différentes étapes qui nous achemineront vers un démonstrateur permettant d'enregistrer et de lire ces nouvelles métadonnées dans une Station de Travail Audionumérique (STAN).

Mots Clés : métadonnées, iXML, *Broadcast Wave Format*, traitements dynamiques, modulation de volume, traitements fréquentiels.

Abstract

As part of a film shoot, the sound material delivered by the sound engineer. is presented in the form of a set of audio files called rushes. In order to facilitate the data exchange between shooting and post-production, these are referenced and intimately linked to additional data, both informative and descriptive. These are called metadata. They provide information on various essential parameters when using the file several months after shooting. Regarding the existing and standardised metadata, there is no information concerning dynamic treatments and filter applied at the time of sound recording that is converted into metadata. The intentions of the sound engineer. are applied directly to the permanently recorded signal and are not referenced as metadata.

However, current technologies are able to carry treatments in a non-destructive manner by using automation curves. As part of this thesis, we will propose new metadata in order to answer various problems related to the non-exploitation of data essential to the soundtrack. The purpose of this research is to present the various steps that will lead us to a demonstrator to record and read these new metadata in a Digital Audio Workstation (DAW).

Keywords: metadata, iXML, *Broadcast Wave Format*, volume modulation, frequency driven dynamics.

Sommaire

Remerciements	3
Résumé.....	4
Abstract	5
Introduction.....	8
Chapitre I - Les métadonnées professionnelles.....	9
I.1/ Historique des fichiers audio	10
I.1.1/ Premiers formats	11
I.1.2/ Fichier Wave	12
I.1.3/ <i>Broadcast Wave Format</i>	15
I.2/ Les métadonnées audio.....	18
I.2.1/ Définition	18
I.2.2/ Premières métadonnées destinées aux professionnels : le BEXT	20
I.2.3/ Seconde génération de métadonnées : le iXML.....	24
I.2.4/ Pérennité des métadonnées audio.....	33
I.3/ Utilisation des métadonnées audio professionnelles existantes	34
I.3.1/ Étape du tournage	34
I.3.2/ Phase de post-production.....	39
Chapitre II – Pourquoi exploiter de nouvelles métadonnées	42
II.1/ Le son direct.....	43
II.1.1/ Importance de la voix dans la bande son	44
II.1.2/ Problématique de l’intelligibilité au cinéma	45
II.2/ Avènement de l’enregistreur multipiste.....	49
II.2.1/ Historique des enregistreurs.....	50
II.2.2/ Pertinence et enjeu du <i>mixdown</i>	53
II.3/ Traitements à la prise	59
II.3.1/ Traitements fréquentiels : le filtrage	60
II.3.2/ Traitements dynamiques	62

Chapitre III – Enregistrement et exploitation de nouvelles métadonnées	71
III.1/ Présentation des nouvelles métadonnées	72
III.1.1/ Définitions.....	73
III.1.2/ Où se situent ces nouvelles métadonnées dans la chaîne du son?	76
III.1.3/ Analyse des paramètres des nouvelles métadonnées.....	77
III.2/ Exploitation de ces nouvelles métadonnées	85
III.2.1/ Mise en forme.....	85
III.2.2/ Question du poids.....	88
III.2.3/ Enjeux liés à l'apparition des nouvelles métadonnées.....	90
III.3/ Implémentation de ces nouvelles métadonnées.....	92
III.3.1/ Présentation des vecteurs d'entrée.....	92
III.3.2/ Phase d'enregistrement.....	94
III.3.3/ Phase de lecture	103
Conclusion	107
Bibliographie.....	108
Table des illustrations.....	113
Table des annexes	115

Introduction

C'est au cours de l'observation de la chaîne du son au cinéma, durant mes différents stages, que la réflexion autour des métadonnées m'est venue. Les métadonnées professionnelles dans le domaine du son constituent une ressource essentielle dans l'exploitation du phénomène sonore après la phase d'enregistrement. À travers la structure RIFF du fichier WAVE, plusieurs tranches décrivant différentes métadonnées ont vu le jour. C'est en premier lieu les métadonnées portant sur les données temporelles qui ont amené à l'apparition de tranches destinées aux professionnelles comme la tranche *Broadcast Extension*. Par la suite, dès 2004, une nouvelle tranche nommée iXML a apporté de nouvelles informations caractérisant plus précisément le contenu des fichiers audio.

D'une part, au regard des informations implémentées dans la tranche iXML, aucune donnée renseignant les traitements tant dynamiques que fréquentiels réalisés par le preneur de son n'est présente sous forme de métadonnée. C'est pourtant ces données qui constituent le cœur de métier dans la prise de son et détermine intrinsèquement la retranscription du son au moment de sa diffusion. D'autre part, la non-exploitation de ces métadonnées engendre différents débats entre les métiers de la chaîne du son notamment autour de la question concernant le *pré* ou *post fader*.

Comment enregistrer ces nouvelles métadonnées renseignant le « geste » du chef opérateur du son au moment de l'enregistrement ? Comment les implémenter dans le cadre d'un flux de travail audionumérique ? Sont-elles exploitables dans le contexte actuel ? À quelles solutions techniques répondent-elles ?

Dans un premier temps, afin de répondre à ces différentes questions, nous analyserons les solutions apportées dans le domaine des métadonnées professionnelles en audio et leur emploi dans la pratique quotidienne d'un travail lié à l'image. Dans un second temps, nous présenterons les méthodes employées lors de la prise de son au cinéma ou à la télévision pour soulever l'ensemble des interrogations que posent les pratiques mises en place. Dans un troisième et dernier temps, nous proposerons des solutions apportées par l'implémentation de nouvelles métadonnées. Les résultats obtenus à travers le développement d'un démonstrateur permettant de les enregistrer et de les lire seront présentés.

Chapitre I

Les métadonnées professionnelles

Les métadonnées sont un support d'information relatif à un contenu préexistant. Dans le domaine de l'audio, ce contenu correspond à toutes les données permettant de décrire le signal numérisé après la Conversion Analogique-Numérique (CAN).

Il faut toutefois garder en mémoire que le monde de l'informatique est scindé en deux entités distinctes : le domaine public et le domaine propriétaire². Ce point est si important qu'il est parfois le moteur d'avancées technologiques mais aussi son propre frein dans la standardisation d'outils informatiques universellement employés dans les différents flux de travail³. Dans un premier temps, nous présenterons les premiers fichiers audio professionnels et plus spécifiquement le fichier WAVE très employé dans le domaine de la prise de son. Dans un second temps, les deux grandes « familles » de métadonnées audio professionnelles seront analysées. Enfin, dans un troisième temps, l'emploi de ces métadonnées dans la chaîne du son sera étudié.

I.1/ Historique des fichiers audio

Pour simplifier la description de l'historique des formats de fichiers⁴ audio, cette partie s'attachera à présenter uniquement les formats les plus importants, sachant qu'il existe de nos jours pratiquement une centaine de formats audio, tant publics que propriétaires. Il faut toutefois retenir que les formats audio actuels sont les héritiers d'une longue gestation.

De la création du premier enregistreur Nagra sur bandes magnétiques par Stefan Kudelski en 1951, alors étudiant à l'école Polytechnique fédérale de Lausanne, au premier enregistreur bipiste Nagra IV-S en 1971, succès dans l'industrie du cinéma, le monde de l'audio a évolué en parallèle des propositions techniques et industrielles. Du son enregistré de manière contiguë sur bandes analogiques aux fichiers audionumériques, il n'y a qu'un pas. La suite des innovations iront dans ce sens avec l'apparition du premier enregistreur sur disque dur, le Deva I de la société Zaxcom.

² Le terme « propriétaire » signifie que les spécifications ne sont pas disponibles dans le domaine public. Il est alors impossible de connaître les méthodes employées de mise en forme et de stockage des données.

³ Le flux de travail, appelé en anglais *workflow*, désigne une chaîne technique qui détermine tous les paramètres en amont du projet, c'est-à-dire avant les étapes de production (le tournage dans le domaine du cinéma) et de post-production.

⁴ Un format de fichier décrit la mise en forme des données dans un fichier informatique.

I.1.1/ Premiers formats

Au début de l'audio numérique, chaque logiciel possédait son propre format de fichier propriétaire. La notion de « format libre » n'est apparue qu'en 1985. Cette même année, l'histoire des fichiers audio débute réellement lorsque la société Electronic Arts développe le format générique *Interchange File Format* (IFF) destiné à la famille des ordinateurs *Amiga 1000*⁵ de la société américaine *Commodore Business Machines Ltd* (CBM). Ces ordinateurs sont très employés, tant par les professionnels que par les particuliers, pour leurs capacités dans le domaine du graphisme et de l'animation. Par exemple, la version grand public de l'*Amiga 1000*, l'ordinateur *Amiga 500*, a été très populaire dans le domaine des jeux vidéo. L'informatique de cette période est marquée par un lien fort entre la mise en forme des données et les puces employées. Par exemple, l'IFF est avant tout adapté à la technologie des processeurs Motorola de la gamme 68000.

L'IFF est un format important dans l'histoire du son car sa structure sous forme de tranches⁶ lui confère un statut particulier, celui d'« ancêtre » le plus proche des formats audio actuels.

Le premier format de fichier audio s'appelait IFF_8SVX. À cette période, de nombreux formats de fichier audio ont vu le jour et déjà les questions autour de la compatibilité entre les formats et de l'échange de données entre les différents logiciels se posaient.

En 1991, les sociétés Microsoft et IBM, s'appuyant sur le format IFF, développent conjointement un modèle de format de fichier spécialisé dans les contenus multimédias, le *Resource Interchange File Format* (RIFF). Ce modèle n'est pas un format de fichier en soi mais il est employé comme architecture originelle des fichiers audio non propriétaires *Audio Interchange File Format* (AIFF) et *Waveform Audio File Format* (WAVE). La différence entre les fichiers AIFF et WAVE était faible car elle reposait principalement sur le sens de lecture du processeur employé :

- **AIFF ;**

Processeurs Motorola avec une orientation *Big Endian*, c'est-à-dire de gauche à droite

- **WAVE.**

Processeurs Intel avec une orientation *Little Endian*, c'est-à-dire de droite à gauche

⁵ Les ordinateurs Amiga sont des micro-ordinateurs créés par la société *Amiga Corporation* et commercialisés par la société *CBM* après 1985.

⁶ Le concept de « tranche » est présenté dans la *Partie I.1.2* du mémoire.

Dans le domaine propriétaire, les fabricants de Station de Travail Audionumérique (STAN) développent leur propre format de fichier. Par exemple la société Digidesign utilise le fichier propriétaire SDII dans Sound Tools devenu depuis Pro Tools. Ce format est fabriqué pour des disques durs formatés⁷ en *Hierarchical File System* (HFS)⁸, formatage utilisé par la société Apple. Il est le fichier audio utilisé au tout début par la société Zaxcom pour encapsuler les données de ses enregistreurs audionumériques. On remarque que les solutions non propriétaires et propriétaires sont employées entre différents fabricants.

Pour simplifier la lecture et l'analyse de l'emploi des métadonnées, nous faisons le choix de ne considérer que le fichier audio WAVE et son évolution, le format *Broadcast Wave* appelé BWF, dans la suite de cette présentation.

I.1.2/ Fichier Wave

Le format WAVE est composé de la manière suivante : d'un *Header*, d'une partie nommée Données (*Data*) et d'un *Footer*. Plus précisément, le WAVE se compose d'un ensemble de données qui sont informatives (les métadonnées) et élémentaires (le son). Les métadonnées sont présentes dans le *Header* et le *Footer* alors que les données du signal numérisé se situent dans la partie *Data*.



Figure 1.1 - Représentation simplifiée du fichier WAVE

L'agencement de ces données est contraint par une mise en forme qui découle de la structure du RIFF. Toutes les données du format WAVE sont divisées en *tranches*, appelées en anglais *chunks*. On peut assimiler le format WAVE à une commode composée d'un nombre indéfini de tiroirs où chaque tiroir correspond à une tranche. Pour identifier une tranche, il suffit de disposer de son nom appelé *descripteur*, de sa taille (nombre d'octets) et de sa mise en forme spécifique employée pour stocker les données.

⁷ Le formatage correspond à l'action de préparer un support de données informatiques (disquettes, disques durs ou encore clés USB) en y inscrivant un système de fichiers normés afin qu'ils soient reconnus par le système d'exploitation de l'ordinateur.

⁸ Le format HFS est un système de fichiers propriétaires développé par Apple pour le système d'exploitation Mac OS. Il a été remplacé par le format HFS+.

Le fichier WAVE possède toujours en en-tête les mêmes informations suivantes en hexadécimal :

- descripteur : 4 octets en codage ASCII⁹ avec une lecture de gauche à droite ;

52 49 46 46 = RIFF

- taille du fichier en octets : 4 octets avec une lecture de droite à gauche ;
- mise en forme : 4 octets en codage ASCII avec une lecture de gauche à droite.

57 41 56 45 = WAVE

Dans sa forme originelle, le fichier WAVE possède deux tranches obligatoires et uniques, la tranche FORMAT (<fmt-ck>) et la tranche DONNÉES (<wave-data>). Toutes les autres tranches sont optionnelles et l'emplacement des tranches dans le fichier n'est pas défini. Toutefois la tranche FORMAT doit se placer avant la tranche DONNÉES afin de connaître certaines informations relatives aux données avant de les lire.

La tranche FORMAT est composée de 16 octets. Elle permet d'indiquer les informations suivantes : codage du son (PCM ou MPEG) sur deux octets, nombre de pistes sur deux octets, fréquence d'échantillonnage sur quatre octets, quantification sur quatre octets, *block align* ou valeur d'alignement¹⁰ sur deux octets et transfert¹¹ (nombre d'octets par seconde) sur deux octets. Elle permet de fournir les informations indispensables à la lecture des données situées dans la tranche DONNÉES. Elles sont indissociables et nécessaire à la lecture du fichier. Le terme « métadonnée » semble alors non approprié pour les décrire.

La tranche DONNÉES se présente de la manière suivante :

- descripteur : DATA ;
- taille ;
- données généralement représentées en *Pulse Code Modulation*¹² (PCM).

À noter que si les données ne sont pas codées en PCM alors une tranche supplémentaire nommée INFORMATIONS (FACT) est obligatoire. Elle apporte d'autres informations sur le contenu du fichier.

⁹ Le terme ASCII correspond à l'*American Standard Code for Information Interchange*. C'est une norme informatique de codage de caractères qui définit 128 codes sur 7 bits, comprenant par exemple des caractères allant des chiffres arabes aux lettres minuscules et capitales. Par exemple, le « a » majuscule correspond à 41 en ASCII.

¹⁰ La valeur d'alignement de bloc en octets est utilisée par un logiciel de lecture pour aligner sa mémoire tampon, appelée *buffer* en anglais. C'est une zone où les informations sont recueillies temporairement lors d'un processus informatique.

¹¹ Le taux de transfert des données permet d'estimer le *buffer* utilisé par le logiciel pour lire sans latence. Ce critère est désormais obsolète au vu de la puissance embarquée dans les ordinateurs actuels.

¹² Le PCM est une technique de quantification numérique d'un signal analogique lors de la CAN. Définie par l'utilisateur, l'amplitude du signal électrique est échantillonnée à des intervalles réguliers.

Lorsqu'un fichier dit polyphonique contient plusieurs signaux numérisés et séparés en différentes pistes, le format de représentation des données en PCM est identique. Néanmoins, les échantillons sont présentés différemment. Par exemple, pour un fichier WAVE contenant deux pistes, les échantillons sont présentés de la manière suivante : l'échantillon n°1 du signal n°1 est suivi de l'échantillon n°1 du signal n°2 qui est suivi de l'échantillon n°2 du signal n°1, etc.

En ce qui concerne le cas des pistes « non armées »¹³, les valeurs des échantillons sont mises à zéro. Cette méthode était employée par Aaton avec le Cantar-X. Désormais, il existe d'autres manières de procéder comme chez Sound Devices où les échantillons des pistes non armées sont supprimés.

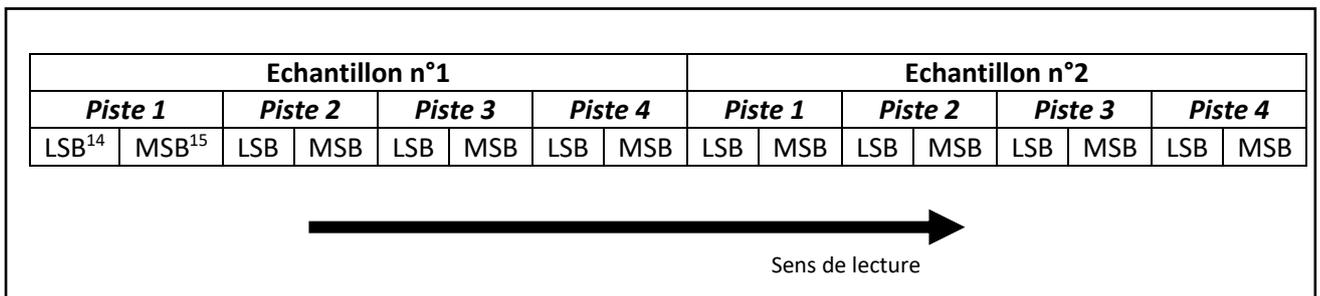


Figure 1.2 - Exemple d'entrelacement d'un fichier polyphonique contenant 4 canaux

Une des particularités du fichier WAVE réside dans sa conception initiale. Sa structure RIFF (sous forme de tranches) lui permet d'être lu par des applicatifs ne reconnaissant pas certaines tranches. Par exemple, lorsque qu'un logiciel de lecture ne reconnaît pas le *descripteur* d'une tranche alors la taille de cette tranche permet d'accéder directement à la tranche suivante.

Dans le fichier WAVE, l'absence de métadonnées temporelles, permettant de relier un fichier audio à un fichier image, rend impossible son utilisation dans un contexte professionnel. Toutefois, grâce à son aspect ouvert et extensible par nature, des améliorations ont été apportées au cours du temps à travers la création de nouvelles tranches (cf. *Annexe A*) et la création de la tranche *Broadcast Extension*.

¹³ Le terme « non armées » désigne les pistes d'un enregistreur qui ne sont pas utilisées lors de l'enregistrement. Par exemple, sur les enregistreurs 24 pistes actuels, une majorité des pistes sont dites non armées car l'utilisateur n'utilise que très rarement l'ensemble des pistes.

¹⁴ Le terme « LSB » signifie *Least Significant Bit* (en français, bit de poids faible). Dans une représentation binaire conventionnelle, il correspond au bit le plus à droite. Dans la pratique, écrit en majuscule, le LSB correspond à l'octet de poids faible.

¹⁵ Le terme « MSB » signifie *Most Significant Bit* (en français, bit de poids fort). Dans une représentation binaire conventionnelle, il correspond au bit le plus à gauche. Dans la pratique, écrit en majuscule, le MSB correspond à l'octet de poids fort.

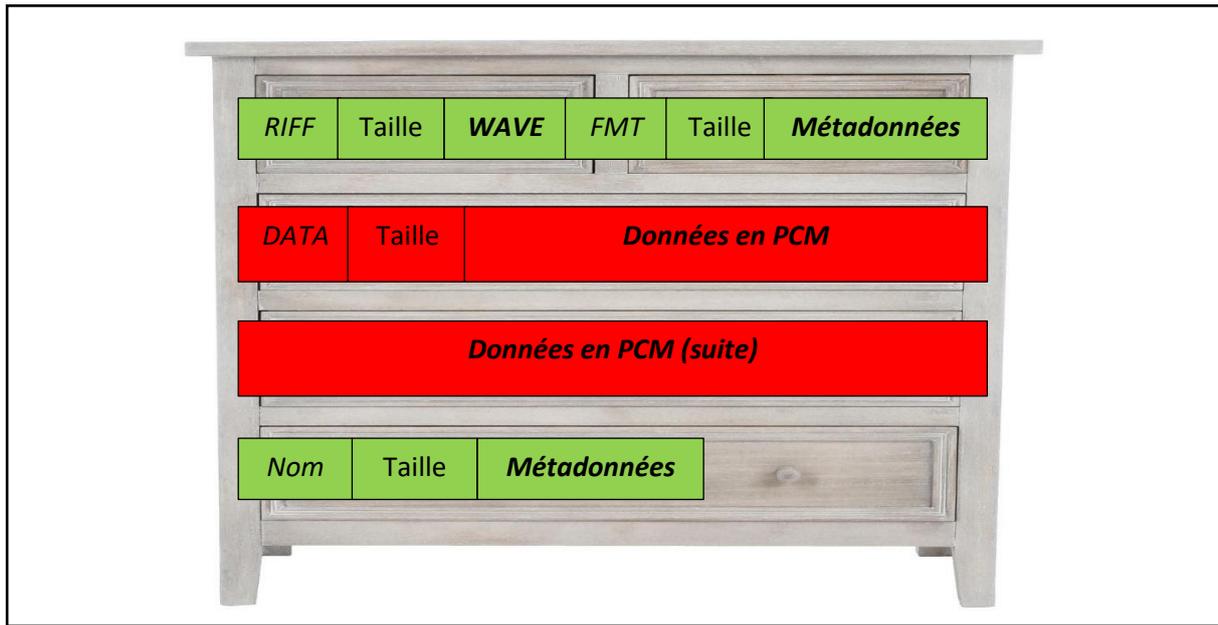


Figure 1.3 - Représentation des tranches historiques du fichier WAVE

I.1.3/ Broadcast Wave Format

De nos jours, pratiquement une centaine de formats audio sont référencés et employés dans différents domaines. Or, dans le contexte de la prise de son, les professionnels utilisent un format de manière presque unanime : le *Broadcast Wave Format* (BWF).

À la fin des années 1990, au moment de la fabrication du premier enregistreur portable audio de la société Aaton, le Cantar-X, la question du format audio a été soulevée. À l'époque, la structure du fichier WAVE non professionnel et grand public présente alors les avantages suivants : format basique, interopérable¹⁶ et évolutif. En créant une nouvelle tranche nommée FILM, Aaton ajoute des métadonnées temporelles essentielles au moment de la post-production. En 1997, revue et modifiée par l'Union Européenne de radio-télévision (UER), la tranche FILM devient la tranche BROADCAST EXTENSION (BEXT) à travers la spécification *Tech 3285 « BWF - A Format for Audio Data Files in*

¹⁶ Le terme interopérable est la « capacité de matériels, de logiciels ou de protocoles différents à fonctionner ensemble et à partager des informations. » [Définition, Dictionnaire Larousse en ligne, 2018 URL : <http://www.larousse.fr/dictionnaires/francais/interop%C3%A9rabilit%C3%A9/43787>]

Broadcasting Version n°0 ». Le BWF est né. Désormais, la version n°2 du même document datant de 2011 et ses six suppléments définissent sa mise en forme normalisée.

En 2003, à la sortie du Cantar-X d'Aaton, le format BWF est implémenté pour la première fois dans un enregistreur audio numérique. Par la suite, en 2004, les sociétés Sound Devices et Zaxcom implémentent à leur tour ce format dans leurs enregistreurs respectifs. C'est ainsi que le format BWF est devenu le format de fichiers audio incontournable dans le monde professionnel, rendu universel par les constructeurs et adoubé par les professionnels. En réalisant un parallèle avec le monde de l'image, les fabricants d'enregistreurs audio ont, dès le milieu des années 2000, implémenté le même format audio alors que les caméras présentent, encore de nos jours, chacune un format différent.

Héritier du fichier WAVE, le BWF reprend l'organisation en tranches. Il possède comme critères communs :

- position non définie des tranches ;

Comme pour le fichier WAVE, les tranches ne possèdent pas d'emplacement prédéfini. Cette caractéristique offre une réelle souplesse et facilite son utilisation. Par exemple, chaque constructeur possède sa manière de mettre en ordre les tranches du BWF. Néanmoins, afin de reconnaître les tranches entre elles, elles sont toutes référencées par un *descripteur* différent, le nom étant normalisé.

- en-tête de fichier commun.

Le BWF possède la même forme canonique avec les mêmes douze octets en début de fichier.

Le BWF comporte obligatoirement les tranches suivantes : la tranche FORMAT, la tranche DATA et la nouvelle tranche BEXT. Comme pour le fichier WAVE, si les données du signal numérisé ne sont pas codées en PCM alors on retrouve, en plus des tranches citées précédemment, la tranche FACT, étendue par rapport au fichier WAVE standard et la tranche MPEG AUDIO EXTENSION. Le BWF reste ainsi rétro compatible avec le fichier WAVE.

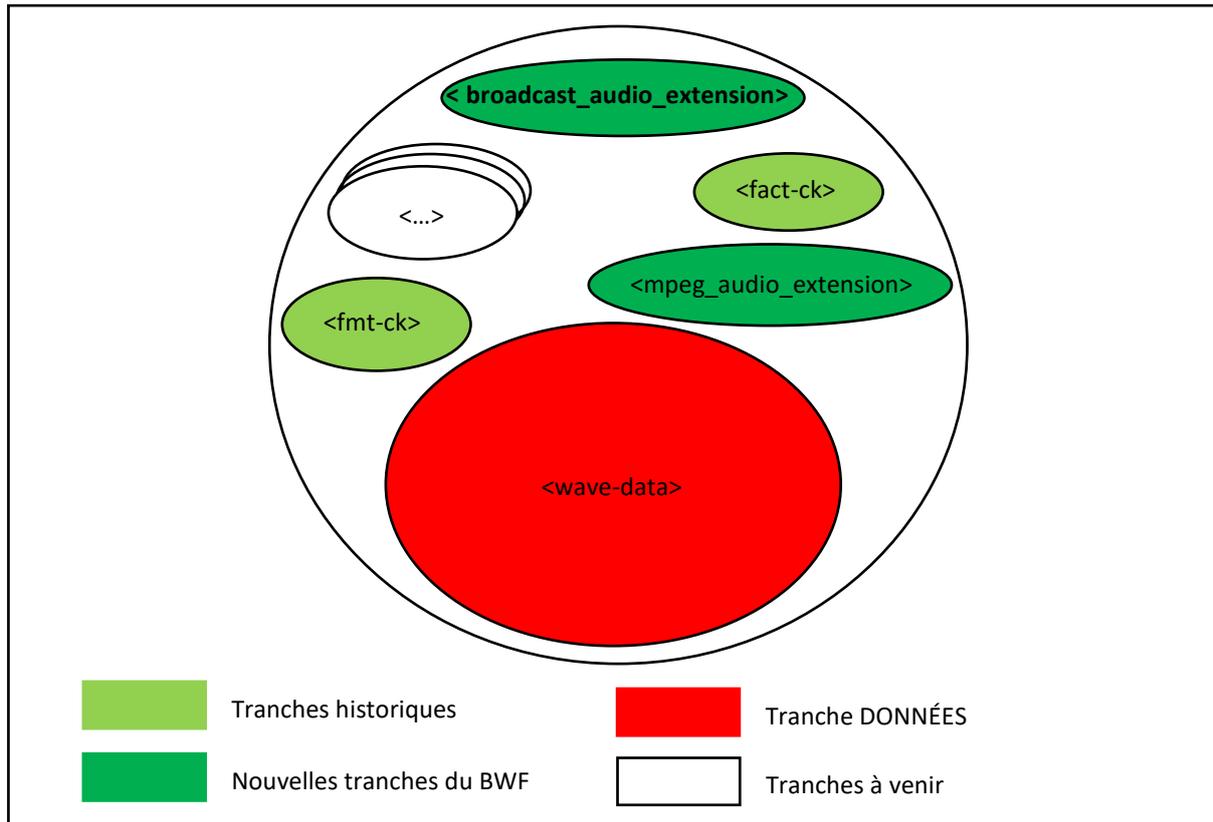


Figure 1.4 - Structure générale du format Broadcast Wave

Au regard des métadonnées apportées par le BWF, on remarque que ce format apparaît comme une version optimisée et adaptée aux professionnels du fichier WAVE. Les modifications apportées se situent principalement dans la nouvelle tranche BEXT qui, contrairement aux tranches historiques, ne sont pas indispensables lors de la lecture des données audio. Elles sont des métadonnées. À ses débuts, le BWF possédait comme extension de fichier .bwf afin de le différencier du fichier WAVE. Depuis, l'extension .wav est utilisé pour ces deux fichiers.

Depuis sa création en 1997, le BWF n'a cessé d'évoluer. À travers les six suppléments rattachés à la spécification *Tech 3285*, de nouvelles tranches facultatives ont été normées. Une liste non exhaustive est présentée en *Annexe A*. Enfin, une version étendue des fichiers WAVE et BWF a été normalisée à travers les formats RF64 et MBWF avec la spécification *Tech 3306* nommée « *MBWF / RF64 : An extended File Format for Audio* » de l'UER. Ces deux formats permettent de dépasser la taille maximum historique du fichier WAVE de 4 gigaoctets grâce à l'emploi d'un codage sur 64 bits et non sur 32 bits. Notons que peu de logiciels se sont alignés sur ces deux formats.

I.2/ Les métadonnées audio

I.2.1/ Définition

Le terme métadonnée est composé du préfixe grec *meta* qui renvoie à la notion de « au-delà de ». Une métadonnée est donc, de manière littérale, une donnée de données.

Elle est une « donnée servant à caractériser une autre donnée, physique ou numérique »¹⁷. C'est une donnée structurée et complémentaire qui permet de caractériser et contextualiser une information dont le sens serait incompréhensible ou inexploitable sans ses métadonnées. Elles sont des outils périphériques permettant de rendre un contenu quelconque intelligible.

La création des premières métadonnées remonte à la gestion des ouvrages dans les bibliothèques et les archives. Afin d'identifier et de contrôler l'ensemble des documents, l'utilisation de fiches cartonnées était obligatoire et même normalisée avec l'*International Standard Bibliographic Description* (ISBD). À l'arrivée d'Internet, le terme « métadonnée » est apparu. On peut référencer deux étapes importantes. La première correspond à un groupe de travail, en 1995 dans l'Ohio, ayant défini un ensemble de descripteurs génériques pour décrire les ressources informatiques du gouvernement américain. Il en résulte la création du *Dublin Core* dont sa structure est normée de manière internationale (ISO 15836). La deuxième étape est liée à l'émergence de langages dits de balisage permettant de fusionner les métadonnées avec leurs fichiers respectifs.

De leur création à leur utilisation, les « *métadonnées sont très diverses : données créées par l'humain ; données créées par la machine ; données destinées à l'humain ; données destinées à la machine.* »¹⁸

De nos jours, la diversité des métadonnées est telle qu'il serait impossible de les uniformiser. C'est pourquoi dans le cadre de ce mémoire, les métadonnées présentées porteront sur deux champs fondateurs de métadonnées audio largement employés par les professionnels. Plus spécifiquement dans le monde du son, l'exploitation des métadonnées se scinde en deux familles : l'une, orientée grand public et l'autre destinée aux professionnels. Les métadonnées orientées grand public sont modifiables à tout moment selon les fonctionnalités offertes par le logiciel. Elles offrent une

¹⁷ Définition, Dictionnaire Larousse en ligne, 2018, URL :

<http://www.larousse.fr/dictionnaires/francais/m%C3%A9tadonn%C3%A9e/186919>

¹⁸ **Jean Debaecker**, De l'usage des métadonnées dans l'objet sonore, Thèse sous la direction de Widad Mustafa El Hadi, Sciences de l'informatique et de la Communication, Université Charles-de-Gaulle –Lille 3, École doctorale S.H.S. - E.D. 473 - Laboratoire Geriico, E.A. 4073, 2012, p.26.

description efficace d'un contenu audiovisuel. Par exemple, iTunes, logiciel gratuit de lecture et de gestion de bibliothèque multimédia, permet à ses usagers de modifier un certain nombre de métadonnées comme le Nom, l'Artiste, le Titre, l'Album, le Compositeur, le Genre, l'Année, le Numéro de pistes ou encore l'illustration du morceau sélectionné.

Actuellement, dans le monde professionnel, les spécifications *Tech 3293* (version 1.8) nommée « *EBU Core Metadata Set (EBUCore)* » de l'UER datant de 2017 et *Tech 3295* (version 2.2) nommée « *P_META - Metadata Library* » de l'UER datant de 2011 fournissent les grandes lignes directrices au niveau de la mise en forme des métadonnées.

Les métadonnées professionnelles se distinguent entre elles de plusieurs manières. En fonction du degré hiérarchique, une métadonnée renseigne un contenu différent. Ce contenu peut désigner un *projet* dans son ensemble, une *prise* spécifique ou une *piste* d'une *prise*. Par exemple, le nom d'un personnage renseigné est défini pour une piste d'une prise. Autre exemple, les métadonnées temporelles implémentées dans la tranche BEXT renvoient quant à elles à des données de prise. Quel que soit le niveau hiérarchique des métadonnées, elles sont toutes divisibles en sous-catégorie avec la gestion de la saisie des informations (manuelles ou automatiques) et dans la manière dont celles-ci varient dans le temps (statiques ou dynamiques).

<i>Hiérarchie</i>	<i>Projet / Piste / Prise</i>			
<i>Nature</i>	<i>Statique</i>		<i>Dynamique</i>	
<i>Gestion de Saisie</i>	<i>Manuelle</i>	<i>Automatique</i>	<i>Manuelle</i>	<i>Automatique</i>

Figure 1.5 - Tableau présentant les différents types de métadonnées professionnelles en audio

Dans le cadre de l'analyse des métadonnées professionnelles, nous considérerons deux tranches du format *Broadcast Wave* : la tranche BEXT et la tranche iXML. Afin de comprendre l'utilité de ces métadonnées et leur emploi au quotidien, des explications ou des mises en situation seront données dans les *Parties 1.2.2* et *1.2.3*.

I.2.2/ Premières métadonnées destinées aux professionnels : le BEXT

Le format *Broadcast Wave* se distingue du fichier WAVE de par l'implémentation d'une nouvelle tranche nommée BROADCAST EXTENSION (BEXT). Répondant une structure RIFF, la tranche BEXT se compose de la manière suivante :

- descripteur : **62 65 78 74** (= BEXT en codage ASCII) ;
- taille : exprimée en octets dont la taille maximum est de 752 octets ;
- données de la tranche BEXT : métadonnées mises en forme comme présentées ci-dessous.

Nom	Taille (en octets)	Format
<i>Description</i>	256 (max)	CHAR (ASCII)
<i>Originator</i>	32 (max)	CHAR
<i>Originator Reference</i>	32 (max)	CHAR
<i>Origination Date</i>	10	CHAR
<i>Origination Time</i>	8	CHAR
<i>Time Reference LOW</i>	4	DWORD
<i>Time Reference HIGH</i>	4	DWORD
<i>Version</i>	2	WORD
<i>UMID</i>	64	BYTE
<i>Loudness Value</i>	2	WORD
<i>Loudness Value</i>	2	WORD
<i>Loudness Range</i>	2	WORD
<i>Max True Peak Level</i>	2	WORD
<i>Max Momentary Loudness</i>	2	WORD
<i>Max Short Term Loudness</i>	2	WORD
<i>Reserved</i>	180	BYTE
<i>Coding History</i>	Non Défini	CHAR

Figure 1.6 - Tableau des métadonnées appartenant de la tranche BROADCAST EXTENSION

L'ensemble des métadonnées de la tranche BEXT sont présentées suivant l'ordre du tableau.

Description

Codé en ASCII, le champ *Description* ne peut dépasser 256 caractères. Il correspond à une description résumée et libre de la séquence. Cette séquence est donc remplie de manière différente selon les fabricants. Par exemple, dans le Cantar d'Aaton, elle comprend les informations suivantes : *Numéro*

de séquence, Numéro de Plan, Numéro de prise, Noms des pistes ou encore Note. Les caractères non utilisés sont remplacés par le caractère nul (0).

Dans la pratique, ce champ a été standardisé par les fabricants de manière officieuse. Les informations essentielles au bon référencement des *rushes*¹⁹ du fichier audio y sont présentes.

Originator

Codé en ASCII, ce champ est composé de 32 caractères maximum. Il permet de référencer le nom de la machine et la version du *firmware*²⁰. Tout comme pour le champ *Description*, les caractères non utilisés sont remplacés par le caractère nul.

Originator Reference

Codé en ASCII, ce champ ne peut dépasser 32 caractères. Il permet de référencer de manière unique la machine employée lors de l'enregistrement grâce à l'*Unique Source Identifier* (USID). La recommandation R099 nommée « *Unique Source Identifier (USID) for use in the <OriginatorReference> field of the Broadcast Wave Format* » de l'UER décrit avec précision cet USID :

CC OOO NNNNNNNNNNNN HHMMSS RRRRRRRRR

- les deux premiers caractères (CC) renseignent le pays en se basant sur la norme *ISO 3166-1* ;
- les trois caractères suivants (OOO) renseignent l'entreprise à partir de la norme *Tech 3279 « International Broadcast Tape Number (IBTN) and Bar-Code Labels »* de l'UER ;
- les douze caractères suivants (NNNNNNNNNNNN) renseignent le nom de la machine et son numéro de série ;
- les six caractères suivants donnent la date au format heure-minute-seconde ;
- les neufs derniers caractères (de 0 à 9) sont renseignés de manière aléatoire par la machine qui utilise un algorithme.

¹⁹ Le terme anglais *rushes* correspond à l'ensemble des fichiers image et audio créés lors d'une production.

²⁰ Le *firmware*, appelé en français micrologiciel ou encore logiciel embarqué, est un programme intégré dans un matériel informatique. Il est nécessaire à son bon fonctionnement.

Cette métadonnée permet de référencer un son de manière intemporelle et unique. L'exemple proposé dans la recommandation R099 est le suivant :

USID : ITRAI DA88396FG347 125324 098748726

D'après l'USID ci-dessus, on peut en déduire que ce fichier a été créé par un Tascam DA88, de numéro de série 396FG347A, utilisé par la RAI (*Radiotelevisione Italiana*) à 12h53mn et 24 secondes.

Origination Date

Codé en ASCII et sur 10 caractères, ce champ renseigne la date de création du fichier audio sur la base du format année (yyyy) - mois (mm) - jour (dd). L'année est définie entre 0000 et 9999, le mois entre 1 et 12 et le jour entre 1 et 31. Chaque terme est séparé par l'un des cinq caractères suivants : « - » *hyphen* ; « _ » *underscore* ; « : » *colon* ; « » *space* et « . » *stop*.

Origination Time

Codé en ASCII sur 8 caractères, ce champ renseigne l'heure à laquelle le fichier a été créé. Le format utilisé est en heure – minute - seconde où chaque élément est composé de deux caractères. Les heures sont bornées entre 0 et 23, les minutes et les secondes entre 0 et 59. Comme pour l'*Origination Date*, chaque terme est séparé par l'un des cinq caractères cités ci-dessus.

Sample Since Midnight / Time Reference

Ce champ renseigne le *Timecode*²¹ codé sur 64 bits en comptant le nombre d'échantillons écoulés depuis minuit. Cette valeur dépend donc de la fréquence d'échantillonnage définie dans la tranche FORMAT. Plus précisément, la valeur donnée indique la position du premier échantillon audio du fichier en comptant les échantillons le séparant de minuit.

Cette référence temporelle est utilisée avec n'importe quelle fréquence d'image et avec n'importe quelle fréquence d'échantillonnage audio. La valeur du *Time Reference* n'est pas donnée directement en *Timecode*. En divisant le *Time Reference* par la fréquence d'échantillonnage (située dans la tranche FORMAT), on obtient, après conversion, le *Timecode* précis du premier échantillon audio enregistré du fichier.

²¹ Le *Timecode* (TC) est une référence temporelle utilisée en audiovisuel. Elle est employée pour la synchronisation des *rushes* image et son enregistrés lors de la phase de production. Le TC s'exprime en heures-minutes-secondes-images.

BWF Version

Codé en éléments binaires non signés, ce champ renseigne la version du BWF implémentée et employée par l'enregistreur. Par exemple, si la version n°1 est utilisée alors le champ *BWF Version* sera égal à 0001h et si la version n°2 est utilisée alors le champ *BWF Version* sera égal à 0002h.

UMID

Ce champ correspond au *Unique Material Identifier*. Les 64 octets sont codés à partir de la norme *SMPTE²² 330M*. Généralement, uniquement 32 octets sont renseignés, on parle alors de « basic UMID ».

Loudness Value

Loudness Range

Max True Peak Level

Max Momentary Loudness

Max Short Term Loudness

Apparues lors de la version n°2 du BWF en 2011, ces métadonnées présentent les différents paramètres de la norme *R128* et des spécifications *Tech 3341* à *3344* de l'UER. Elles renseignent sur les critères fondamentaux permettant de caractériser le niveau d'intensité moyen d'un programme audiovisuel.

Reserved

Les 180 octets sont réservés à des extensions futures de la norme. Si le champ *BWF Version* est égal à 0001h ou 0002h alors l'ensemble de ses octets sont mis à 0.

Coding History

Codé en ASCII, ce champ permet de décrire un historique des processus de codage appliqués à l'audio. Afin de simplifier sa lecture, l'UER a fourni une recommandation, la spécification *R 98* en 1999, permettant de normaliser ce champ.

²² L'acronyme « SMPTE » désigne la *Society of Motion Picture and Television Engineers*, association internationale fondée en 1916 et située aux États-Unis. Elle développe des standards et des spécifications employés dans le domaine de la télévision et du cinéma numérique.

Au regard des métadonnées proposées dans la tranche BEXT, on peut remarquer qu'elles correspondent toutes, dans le cadre d'un travail à l'image, à des informatives élémentaires. Afin de relier un fichier image et un fichier son, il faut obligatoirement disposer de données temporelles. Comme pour la bonne lecture d'un fichier WAVE ou BWF avec la tranche FORMAT, la tranche BEXT est obligatoire. La tranche BEXT est donc un complément indispensable au fichier WAVE originel qui souffrait de l'absence de métadonnées temporelles.

Les métadonnées autres que temporelles, qui informent sur le contenu sonore ne sont, quant à elles, peu voire mal spécifiées. Le champ *Description* de la tranche BEXT est le plus utile mais est décrit de manière libre, il varie donc d'un constructeur à l'autre malgré un effort d'homogénéité de la part des fabricants. C'est pourtant dans ce champ spécifique que se regroupent de nombreuses informations utiles aux métiers de la post-production.

C'est à travers une demande de plus en plus forte de la part des utilisateurs et des constructeurs que les informations contenues dans le champ *Description* ont été mises en forme. La tranche iXML est née de ce constat.

1.2.3/ Seconde génération de métadonnées : le iXML

Le iXML est apparu le 8 Juillet 2004 lors d'une convention entre fabricants audio organisée par l'*Institute of Broadcast Sound (IBS)* à Londres. Le « i » de iXML fait référence au rôle majeur joué par l'IBS dans le fait d'avoir rassemblé différents constructeurs pour résoudre les enjeux des métadonnées en audio. En évoquant les difficultés d'évolution de la tranche <BEXT> du format BWF, Mark Gilbert de la société UK Gallery propose et documente une nouvelle tranche où la description des métadonnées est simple, souple et extrêmement évolutive. Il propose alors le iXML, un langage *open source*²³ basé sur le XML²⁴. Le choix d'utiliser le XML tient dans le fait que ce langage possède un code compatible avec les procédés informatiques existants, facile à implémenter afin d'être rapidement adopté. Bien que les fichiers audio aient déjà des données temporelles, la proposition est acceptée de par la souplesse du format.

²³ Le terme *Open Source* désigne les logiciels ou applicatifs dont le code source est disponible dans le domaine public.

²⁴ Le terme XML renvoie au langage informatique de balisage *Extensible Markup Language* dérivant du langage SGML. Encodé en UTF-8, le XML permet d'utiliser de nombreux caractères indisponibles dans le codage ASCII.

Un groupe de travail est organisé entre les sociétés suivantes afin de présenter une version mature : Aaton, Apple, Ambient, Avid, BBC, Dark Matter, Digital, Digidesign, Fairlight, Fostex, Gallery, HHB, Lightworks, Merging technologies, Nagra, SadiE, Sound Devices, Sound of Voices, Steinberg, Synchronarts, Videosonics et Zaxcom. Quelques mois plus tard, la première version officielle du iXML voit le jour. La première machine à implémenter la tranche iXML est le *Metacorder* de la société Gallery UK.

Désormais, la tranche iXML est inscrite dans le fichier BWF. Elle n'est pas normée (non officielle au niveau des instances internationales) ainsi sa forme actuelle est administrée par la société privée Gallery UK et son créateur Mark Gilbert à travers le site de référence du iXML²⁵ régulièrement mis à jour. La « spécification » du format est donc libre et évolue au fil des années au cours de discussions entre utilisateurs, fabricants d'enregistreurs et de STAN. Par exemple, la tranche iXML pèse désormais 61 kilooctets alors que sa première version dans le Cantar-X ne faisait que 8 kilo-octets. Les constructeurs implémentant le iXML dans leurs machines fournissent une charte d'implémentation permettant d'indiquer les métadonnées écrites et lues (cf. *Annexe B* présentant la charte d'implémentation iXML de l'enregistreur SX-R4+ de Sonosax).

À travers l'approbation de nombreux fabricants audio, le iXML est définitivement devenu un standard lorsque la société Apple a introduit le iXML dans le logiciel de montage vidéo *Final Cut Pro* avec la version 6.0.2 en fin 2007.

En 2014, lors de l'*International Broadcasting Convention* (IBC), salon destiné aux professionnels du divertissement et de la communication se déroulant tous les ans à Amsterdam, l'Association Française du Son à l'Image (AFSI) a proposé de nouvelles métadonnées alors non exploitées dans cette tranche. Cette proposition est liée à la généralisation des enregistreurs multipistes en tournage où la demande de rapport son numérique est de plus en plus en forte. Après adoption de certaines de ces propositions par Aaton et Sound Devices dans leurs enregistreurs respectifs, ces nouvelles balises ont été validées en partie par Gallery UK en Juillet 2016 et sont apparues dans la version 2.01 du iXML.

Au niveau de sa structure, le iXML utilise des balises appelées *tags*. Trois types de balises sont employées : booléenne (vrai ou faux), dictionnaire ou libre. Au niveau de la mise en forme du iXML, sa structure la plus simple est présentée en (1) de la *Figure 1.7*. On retrouve aussi des structures

²⁵ Consultable à l'adresse URL suivante : <http://www.ixml.info/>

imbriquées ou dites de *groupe* présentées en (2) voire des balises *master* en (3) permettant de créer des ensembles de groupe.

<code>< Balise > : début de balise</code> Texte libre <code></Balise> : fin de balise</code>	<code><Balise_Groupe></code> <code> < Balise ></code> <code> </Balise></code> <code></Balise_Groupe></code>	<code><Balise_Master ></code> <code> <Balise_Groupe></code> <code> < Balise ></code> <code> </Balise></code> <code> </Balise_Groupe></code> <code></Balise_Master></code>
(1)	(2)	(3)

Figure 1.7 - Tableau des différentes balises dans la tranche iXML

Comme présenté en (1) de la Figure 1.7, le texte est encadré de deux balises XML, une ouvrante et l'autre fermante. Cette structure de base définit un *paramètre*. Un *paramètre* peut être encadré de deux balises ce qui forme un *conteneur d'objets* (2). Enfin, un *conteneur d'objets* peut être regroupé dans un autre *conteneur d'objet* comme présenté en (3).

La tranche iXML est une tranche de type RIFF, elle se présente de la manière suivante :

- descripteur : **69 58 4D 4C** (= iXML en codage UTF-8) ;
- taille : exprimée sur quatre octets ;
- données de la tranche iXML présentées ci-dessous.

Comme pour la tranche BEXT exposée en *Partie 1.2.2*, nous décrivons les métadonnées principales inscrites dans le iXML en suivant l'ordre des balises standardisées par Mark Gilbert dans la version 2.02 présentée en *Annexe C*. Elles sont divisibles en dix groupes distincts. Afin de visualiser et de différencier plus facilement les types de balises, un code couleur est utilisé :

<Balise_Master>
<Balise_Groupe>
<Balise>
<Balise_En cours de développement>

<BWFXML>

Cette première balise est obligatoire car elle permet d'indiquer le début de la tranche iXML.

Groupe n°1 : les métadonnées principales

Ce premier groupement présente 14 balises de la tranche iXML. Il correspond à une mise en forme « normalisée » de l'ensemble des métadonnées présentes dans le champ *Description* dans la tranche BEXT.

<IXML_VERSION>

Cette balise permet de renseigner la version implémentée dans l'enregistreur et de connaître les métadonnées renseignées par l'enregistreur.

<PROJECT>

Cette balise permet de référencer le nom du projet. La limitation de caractères dépend des fabricants.

<SCENE>

<TAPE>

<TAKE>

Ces trois métadonnées permettent de relier une image donnée à un fichier son enregistré séparément à travers le numéro de scène et de prise. La *TAPE* correspond au « numéro de cassette ». Par défaut, elle est automatiquement donnée de la manière suivante mois (mm) - jour (dd). Il est vivement conseillé de lui donner comme valeur la date de création du fichier car la société Avid utilise cette métadonnée afin de distinguer deux fichiers enregistrés à la même heure mais sur deux journées différentes.

Ces métadonnées sont extrêmement importantes car elles facilitent la recherche dans les STAN lors de la phase de post-production. Elles permettent de réaliser des regroupements ou familles de fichiers correspondant à une même scène tournée à l'image. Par exemple, si un monteur son désire modifier un mot ou remplacer une phrase entière d'un comédien, il utilise ces trois métadonnées pour rechercher les fichiers ayant enregistrés la même scène et par conséquent les mêmes répliques.

<TAKE_TYPE>

Introduite dans la version 2.0, cette métadonnée utilise un dictionnaire qui permet de définir la prise enregistrée. Les mots disponibles aux utilisateurs sont les suivants : DEFAULT, NO_GOOD, FALSE_START, WILD_TRACK, PICKUP, REHEARSAL, ANNOUNCEMENT et SOUND_GUIDE. Cette balise permet une meilleure gestion des fichiers et elle indique clairement la fonction du son enregistré.

Le terme IMPULSE_RESPONSE et l'emploi de mot n'appartenant pas à un dictionnaire prédéfini ont été proposés par l'AFSI en 2014 mais n'ont pas été acceptés.

<NO_GOOD>

<FALSE_START>

<WILD_TRACK>

Ces trois métadonnées sont désormais obsolètes depuis la version 2.0 avec l'évolution de la métadonnée TAKE_TYPE présentée ci-dessus.

<CIRCLED>

Renseignée sous forme de valeur booléenne, c'est-à-dire 0 ou 1, elle permet de repérer une bonne prise.

<FILE_UID>

C'est un nombre qui permet d'identifier de manière unique le fichier partout dans le monde. Ce chiffre reprend le nom du fabricant, le nom de la machine, le numéro de série de la machine, la date, l'heure, trois chiffres donnés au hasard et quatre caractères indiquant le numéro de la piste.

<UBITS>

La métadonnée *UBITS* est générée à partir de différents procédés. Elle peut être issue du *Timecode* ou de la date de création. Elle est très rarement utilisée à cause de nouvelles métadonnées qui remplissent cette fonction.

<NOTE>

La balise *NOTE* permet de contenir un commentaire réalisé par l'ingénieur du son au moment de la prise. Elle peut renseigner des défauts, des commentaires sur les voix des acteurs ou encore indiquer, à une valeur temporelle, un événement sonore intéressant ou indésirable. Elle permet de donner une vision d'ensemble de la prise et de transporter, à la phase de post-production, un sentiment ou une émotion. Il n'y a pas de limitation de caractères dans la « norme », qui dépend uniquement des fabricants.

Selon les fabricants, l'emploi de la balise *NOTE* est différent. Par exemple, dans le Cantar-X3 d'Aaton, on retrouve des « notes permanentes » (*persistent notes*) qui sont automatiquement inscrites dans la balise *NOTE* de toutes les prises suivantes jusqu'à son effacement.

<PRE_RECORD_SAMPLECOUNT>

La balise *PRE_RECORD_SAMPLECOUNT* permet de communiquer la valeur du *buffer* de pré-enregistrement. La valeur de cette balise est donnée en échantillons. Depuis la version 1.4 et l'apparition de la balise *SYNC_POINT_FUNCTION*, la balise *PRE_RECORD_SAMPLECOUNT* est devenue obsolète.

D'autres métadonnées sont disponibles dans la spécification du iXML, une description simple est donnée ci-dessous. Pour plus de détails, des informations supplémentaires sont disponibles en *Annexe C*.

Groupe n°2 : les points de synchronisation

<SYNC_POINT_LIST>

Ce groupe contient plusieurs métadonnées qui permettent d'indiquer des points de synchronisation présents dans le fichier.

Groupe n°3 : la fréquence d'image spécifique à une prise

<SPEED>

Le groupe *SPEED* est utilisée pour les *rushes* non synchrones, c'est-à-dire où la fréquence d'image est modifiée pour un plan spécifique. Par exemple, les images au ralenti emploient des fréquences d'image différentes des plans courants dans un souci de traitement en post-production. Ce groupe comporte 11 balises.

Groupe n°4 : les paramètres de la norme R128

<LOUDNESS>

Le groupe *LOUDNESS* reprend les mêmes valeurs indiquées dans la tranche BEXT du fichier BWF. Ce groupe est uniquement une recopie d'informations.

Groupe n°5 : le nom des fichiers

<HISTORY>

Le groupe *HISTORY* permet de créer un lien de parenté entre différents fichiers enregistrés simultanément.

Groupe n°6 : les métadonnées de conformation²⁶

<FILE_SET>

Le groupe *FILE_SET* est composé de quatre métadonnées qui permettent de faciliter la conformation automatiques des *rushes* de tournage.

Groupe n°7 : les métadonnées de piste

<TRACK_LIST>

Ce groupe indique différentes informatiques spécifiques à chaque piste enregistrée. En 2014 lors de l'IBC, l'AFSI a proposé de nouvelles sous-balises afin de référencer avec une plus grande précision une piste. Ces nouvelles métadonnées concernent essentiellement des informations sur les traitements réalisés lors de la prise de son. Voici les neuf balises proposées par l'AFSI.

<TRACK_TYPE>

Cette balise utiliserait un dictionnaire avec les mots suivants : DEFAULT, WILD_SOUND, PLAYBACK, SOUND_GUIDE, NO_SOUND. Elle a été acceptée par les fabricants mais elle ne figure pas dans la dernière version du iXML.

À travers le terme NO_SOUND, cette balise pourrait faciliter le processus d'*auto-conformation* en éliminant les fichiers muets. Le terme WILD_SOUND pourrait quant à lui clarifier la fonction des pistes servant d' « ambiance-raccord » au moment d'une prise.

²⁶ Au moment où le montage est capable de délivrer une version aboutie du montage, par le biais d'un super-conteneur comme un AAF, la session de montage est transmise au monteur paroles. À partir du montage son réalisé durant le montage image, les pistes divergées des *rushes* son doivent être développées : c'est l'étape de conformation, appelée auto-conformation des *rushes* son.

<INPUT_COUNT>

Cette balise indiquerait le nombre d'entrées pour cette piste.

<INPUT_SETTING>

Cette balise contiendrait le nom de la marque des microphones employés.

<PHASE>

Cette balise permettrait d'indiquer l'emploi ou non de l'*opposition de phase* du signal.

<DELAY> / <INPUT_DELAY>

De par l'emploi de plus en plus fréquent d'émetteur-récepteur numérique pour les microphones cravates²⁷, les enregistreurs actuels offrent la possibilité de modifier le *delay* en entrée de piste. Présente dans la balise privée d'Aaton du iXML, sa valeur s'exprimerait en nombre d'échantillons.

<INPUT_FILTER>

Cette balise permettrait de connaître le filtrage coupe-bas analogique employé lors de la prise. Elle existe déjà dans la balise iXML privée d'Aaton.

<EQ>

Cette balise permettrait de connaître l'égalisation employée lors de la prise.

<TRACK_SOURCE>

Cette balise donnerait une indication sur la fabrication du *mixdown*²⁸.

<TRACK_NOTE>

Contrairement à la balise *Note*, cette balise offrirait la possibilité d'écrire des commentaires spécifiques à une piste dédiée.

²⁷ Le microphone cravate est l'évolution du microphone « Lavallier » que l'on portait en collier sur les plateaux. Ce type de microphone est basé sur la technologie des microphones à électret. On parle aussi de microphone miniature car ils sont facilement dissimulable dans les vêtements et les chapeaux. Généralement, on emploie le terme de « microphone HF » pour les désigner.

²⁸ Le terme *mixdown* correspond à une réduction de l'ensemble des pistes en deux pistes. Il est défini et expliqué en *Partie II.2.2*.

Groupe n°8 : les métadonnées du BEXT

<BEXT>

Le groupe *BEXT* reprend à l'identique les informations contenues dans la tranche BEXT du format BWF et constitue donc une copie de la tranche BEXT dans la tranche iXML. L'apparition de cette balise suppose une suppression de la tranche historique BEXT dans les années à venir.

Groupe n°9 : les métadonnées pour les coordonnées personnelles

<USER>

Le groupe *USER* se compose de quinze balises permettant de renseigner les coordonnées des différents intervenants de la chaîne du son afin de simplifier l'étape de post-production et de ne plus utiliser de rapport son papier. En 2013, l'AFSI demandait déjà ces métadonnées dans le iXML ainsi le travail de cette association a permis la création de cette balise.

Groupe n°10 : la localisation

<LOCATION>

Le groupe *LOCATION* date de la version 2.0 et comprend cinq balises indiquant les coordonnées géographiques d'un lieu.

Ces métadonnées sont utiles pour savoir si deux plans d'une même séquence ont été tournés au même endroit. Elles permettent de faciliter la compréhension des *rushes* son.

Dans l'*Annexe E* est décrit l'ensemble des balises du iXML en reprenant le niveau hiérarchique, la gestion de saisie et la manière dont elles varient dans le temps.

L'ensemble des balises présentées ci-dessus correspondent aux métadonnées du domaine public. La « norme » iXML n'exclut pas d'écrire d'autres groupes et balises privées, utilisées par les fabricants afin de décrire d'autres métadonnées. Il existe donc une zone définie par la « norme » qui regroupe l'ensemble des balises décrites et une zone privée destinée au fabricant. Généralement, cette zone privée regroupe des informations n'ayant pas été implémentées de la même manière selon les constructeurs et dont le fonctionnement est propre à la machine. Par exemple, Aaton utilise la balise *AATON_CANTAR* pour décrire de nombreuses métadonnées et conserver une trace de la configuration de l'enregistreur au moment de l'acquisition.

La spécification du iXML est en constante évolution. Au regard des métadonnées présentes dans cette tranche, on remarque que l'ensemble des informations est utile au moment de la lecture des fichiers en post-production, lors du montage paroles. Ces métadonnées correspondent à une demande spécifique dans la chaîne du son à l'étape du tournage, elles ne sont pas employées dans l'archivage des fichiers.

I.2.4/ Pérennité des métadonnées audio

Depuis la création de la tranche iXML (2004), de nouveaux formats d'échange de contenu multimédia ont vu le jour comme le *Material eXchange Format* (MXF). Ce super conteneur défini par la *Society of Motion Picture & Television Engineers* (SMPTE) dans la spécification « *ST 377-1:2011 - Material Exchange Format (MXF) — File Format Specification* » propose de véhiculer les informations d'image et de son dans un même et unique fichier. Le fichier BWF, contenant uniquement du son, est utilisé comme une essence dans le MXF, symbole de sa pérennité dans le temps. On retrouve aussi le BWF employé dans d'autres supers conteneurs comme l'*Advanced Authoring Format* (AAF) et l'AES31 développé par l'*Audio Engineering Society* (AES). De plus, le fichier BWF est le format par défaut dans la STAN Pro Tools d'Avid, standard dans l'industrie du cinéma et de la musique, depuis la version n°5. On retrouve aussi le fichier BWF jusqu'à la phase d'exploitation en cinéma, avec la copie de distribution numérique, en anglais appelée *Digital Cinema Package* (DCP).

Développer de nouvelles tranches dans le fichier BWF ne représente qu'une des premières étapes. Avant d'être adoptée par l'ensemble des métiers de la chaîne du son, son développement doit faire partie d'un véritable partenariat entre deux entités situées à l'opposé de la chaîne. Par exemple, la création de la tranche FILM a été possible grâce au partenariat entre les sociétés SADiE et Aaton. Malgré l'effort réalisé par le domaine du tournage et de la prise de son pour implémenter et standardiser les métadonnées de la tranche iXML, la difficulté est de les lire dans les STAN en post-production. Par exemple, jusqu'à la version 11 de Pro Tools, les tranches de métadonnées non reconnues étaient automatiquement supprimées.

Ces dernières années, une des évolutions les plus importantes dans le domaine des métadonnées professionnelles en audio est l'abandon de la tranche <axml> standardisée par l'UER en faveur de la tranche iXML, non reconnue par les instances de normalisation (UER, AES et SMPTE).

À l'heure actuelle, la tranche iXML reste non officielle mais elle est implémentée par la grande majorité des constructeurs en audio. Elle représente une « norme officieuse ». Au niveau de sa reconnaissance officielle, l'AES a enclenché un groupe de travail en mai 2008, l'AES-X155 intitulé « *Production recording metadata set (iXML)* ». En février 2017, dans le compte-rendu de la réunion du département son de la Commission Supérieure Technique de l'Image et du Son (CST), il est demandé de relayer les travaux réalisés par l'AFSI afin de tendre vers une proposition de standardisation de la tranche iXML par l'AES.

I.3/ Utilisation des métadonnées audio professionnelles existantes

Les métadonnées professionnelles en audio sont créées au moment de la prise de son et exploitées par la suite en post-production. Cette partie présentera les outils employés par les professionnels pour les manipuler durant les différentes phases de la chaîne du son.

I.3.1/ Étape du tournage

L'étape d'enregistrement représente la phase la plus importante pour les métadonnées. C'est au moment de la création des fichiers que les métadonnées sont enregistrées dans les différentes tranches comme la tranche BEXT ou iXML. Non renseignées, elles ne sont pas exploitables par la suite et rendent difficile la gestion du fichier au moment de la post-production. Comme présenté dans l'Annexe E, elles sont inscrites soit par la machine, c'est-à-dire de manière automatique par l'enregistreur, soit manuellement par le chef opérateur du son²⁹.

Pour améliorer la gestion des métadonnées, les constructeurs développent désormais des applicatifs sur tablette et *smartphone* déportés des enregistreurs contrôlant les machines à distance via le *Bluetooth* ou le *Wi-Fi*. Par exemple, la société Zoom a développé le *F8 Control* contrôlant l'enregistreur *F8* à distance. On retrouve les applicatifs similaires avec *Wingman* chez Sound Devices, *Nomad Touch* chez Zaxcom et les applicatifs web des enregistreurs *SX-R4+* de chez Sonosax et *Cantar-X3* d'Aaton. Ces outils permettent à l'utilisateur de contrôler le transport, la gestion des métadonnées et les

²⁹ Le terme « chef opérateur du son », « ingénieur du son » ou encore « opérateur du son » sont employés indifféremment et désigne généralement le même métier dans le cadre du cinéma.

différents traitements comme le panoramique. Néanmoins, l'usage de tablette ou *smartphone* peut ne pas être adapté aux conditions extrêmes de tournage à cause des températures et de l'humidité.

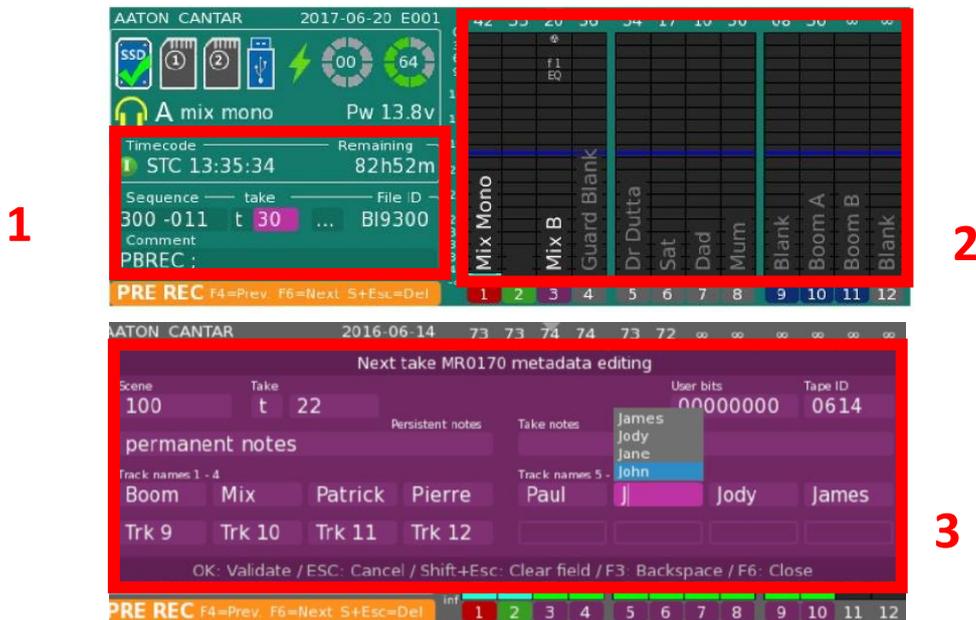


Figure 1.8 - Capture d'écran du Cantar X-3 (version 2.23)

L'interface du Cantar X-3 permet de présenter avec une grande clarté la gestion des métadonnées pendant l'enregistrement. Sur les deux images ci-dessus, dans les encadrés n°1 et n°2, on observe que l'utilisateur possède un contrôle visuel des métadonnées de *BWF_TIME_REFERENCE_LOW* et *BWF_TIME_REFERENCE_HIGH*, *SCENE*, *TAKE*, *COMMENT* et *NAME*. Dans l'encadré n°3, il peut modifier l'ensemble de ces métadonnées mis à part les métadonnées temporelles. Il apparaît clairement que les métadonnées les plus importantes pour la post-production sont directement accessibles au moment de la prise de son. Les entretiens réalisés avec des monteurs paroles et des monteurs son ont confirmé que la saisie des noms donnés aux différentes pistes était primordiale lors de l'enregistrement.

Quant aux autres balises de la tranche iXML, elles sont renseignées dans d'autres sections de l'enregistreur. Cependant, se pose la question du remplissage de tous les champs et l'emploi qui en est fait en post-production. Est-il nécessaire de renseigner un fichier dont les métadonnées ne seront pas lues en phase de montage ?

En parallèle de la gestion des métadonnées dans l'enregistreur, le chef opérateur du son rédige un rapport son papier. Cette « base de données » manuscrite permet de renseigner les informations du tournage en deux temps. Dans la partie supérieure, on retrouve des informations nécessaires à la bonne gestion des *rushes*. Elle informe sur les paramètres suivants :

- | | |
|-------------------------------|----------------------------------|
| - date ; | - fréquence d'échantillonnage ; |
| - nom de Production ; | - quantification ; |
| - titre du Film ; | - type de fichiers enregistrés ; |
| - nom de l'ingénieur du son ; | - format de sauvegarde. |
| - fréquence d'image ; | |

Au cœur du document (*cf. Figure 1.9*) se trouvent les informations propres aux différentes prises. Ces paramètres se présentent de la manière suivante :

- | | |
|--------------------------|---------------------|
| - nom du fichier ou ID ; | - noms des pistes ; |
| - séquence et Plan ; | - commentaires. |
| - numéro de Prise ; | |

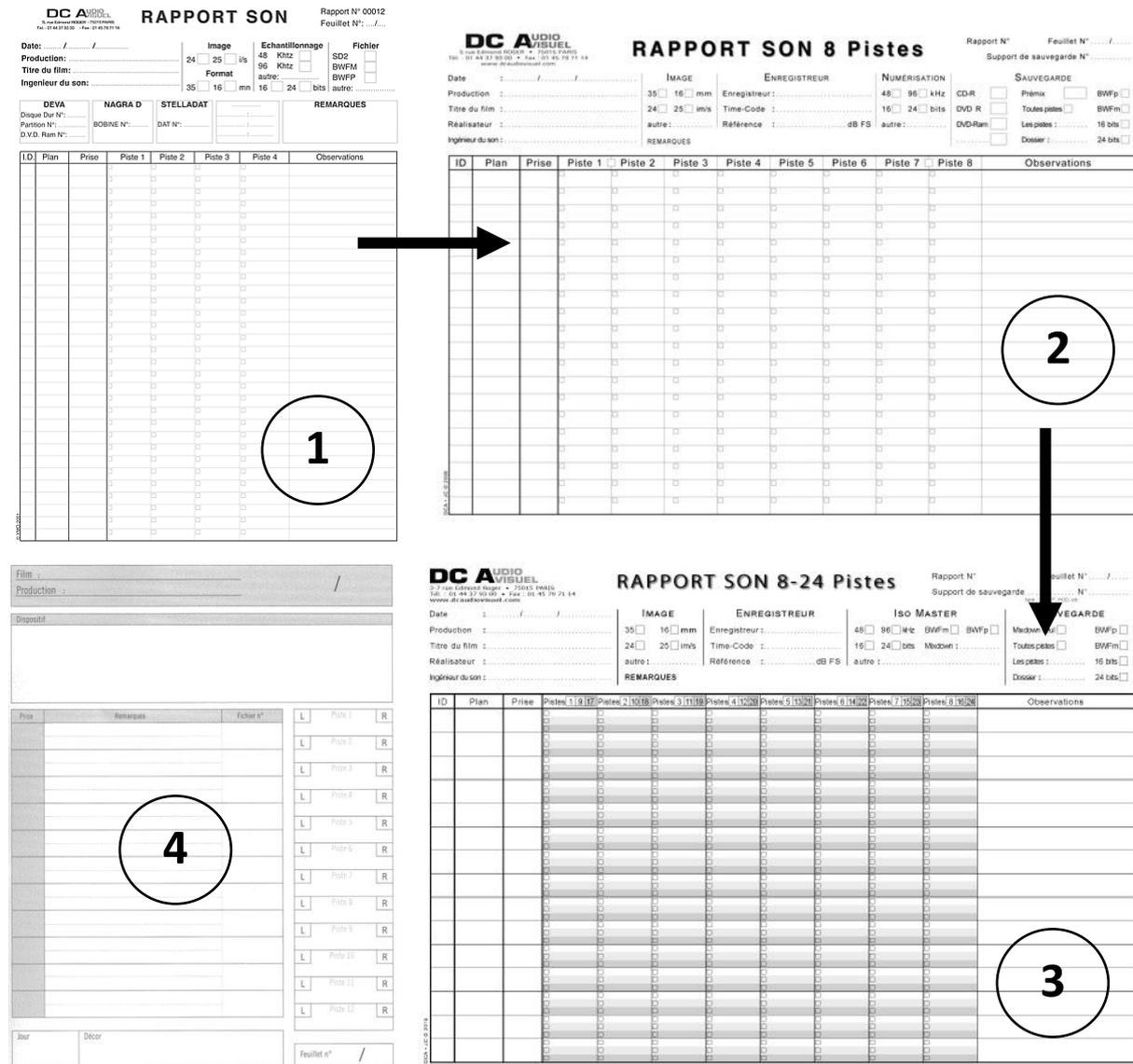


Figure 1.9 - Présentation des différents rapports son papier
 (1 – Rapport sur 4 pistes de DCAudiovisuel ; 2 – Rapport sur 8 pistes de DCAudiovisuel ;
 3 – Rapport sur 24 pistes de DCAudiovisuel ; 4 – Rapport sur 24 pistes de Tapages & Nocturnes)

L'évolution du rapport son (cf. Figure 1.9) est intimement liée à l'avènement des enregistreurs multipistes et notamment des enregistreurs 24 pistes (2015). On peut aussi remarquer que l'ensemble des données retranscrites dans le rapport son sont désormais toutes disponibles sous forme de métadonnées dans la tranche iXML. De plus, les feuilles de rapport son qui accompagnaient les bobines magnétiques ou les cassettes DAT tendent à être remplacées par des rapports générés automatiquement par les machines. Par exemple, avec le Cantar-X3, l'utilisateur peut générer un fichier au format *Portable Document Format* (PDF), se présentant sous la forme d'un tableau et reprenant l'organisation des rapports son manuscrits.

Toutefois dans la section *Commentaire*, le rapport son manuscrit permet de témoigner avec une grande précision de l'ensemble des « incidents » sonores (avion, circulation, bruits de perche ou encore frottements d'un microphone cravate) survenus lors de la prise et sur l'intelligibilité des dialogues. Par exemple, le chef opérateur du son Lucien Balibar utilise un système de couleur afin de classer par ordre d'importance les commentaires qu'il réalise :

*« Les appréciations doivent enfin faire ressortir visuellement les éléments utiles ou remarquables, et à l'inverse les problèmes mineurs ou majeurs : j'ai choisi pour ma part d'utiliser un code simple, avec un stylo bille quatre couleurs ; y figure en vert tout ce qui doit être écouté avec attention par le monteur, et en rouge tout ce qui m'a posé problème ou peut en poser lors de l'exploitation des sons directs. »*³⁰
(Lucien Balibar)

Lors de l'atelier « Le son direct »³¹ organisé par l'AFSI en 2013, la question du rapport son a été soulevée. Il est écrit que la meilleure gestion des métadonnées par les logiciels de post-production facilite la recherche de fichiers et rend le rapport son papier « accessoire ». De plus, tenir à jour en parallèle le renseignement des métadonnées dans l'enregistreur et une version papier semble une charge de travail un peu disproportionnée au vu des informations redondantes entre les deux supports. La question reste d'actualité puisque récemment, au Micro Salon AFC de 2017, un atelier portant sur l'avenir du rapport son a été organisé par l'AFSI.

En plus des métadonnées et du rapport son manuscrit ou numérique, le chef opérateur du son a la possibilité de générer d'autres fichiers informatiques. On retrouve trois types de tableaux éditables qui récapitulent toutes les métadonnées d'un fichier audio :

- Fichier .tsv (*Tab-Separated Values*) ;

Fichier tableur contenant des données séparées par des tabulations pour les colonnes et des retours à la ligne pour les lignes.

- Fichier .csv (*Comma-Separated Values*) ;

Fichier tableur contenant des données séparées par des virgules pour les colonnes et des retours à la ligne pour les lignes. Il est très employé pour générer les rapports son de tournage.

³⁰ **Lucien Balibar**, *La chaîne du son au cinéma et à la télévision : De la prise de son à la post-production*, Paris, Dunod, coll. Audio-Photo-Vidéo, 2015, p. 162.

³¹ **Thierry Lebon**, « Compte-rendu - L'atelier « Le son direct » », Association Française du Son à l'Image, juin 2013, URL : <https://www.afsi.eu/articles/15233-atelier-le-son-direct>

- Fichier .ALE (*Avid Log File Format*).

Fichier utilisé par la société Avid dans ses logiciels d'édition vidéo. Les métadonnées sont exportées dans un format spécifique afin de les rendre compatibles avec les logiciels d'Avid.

L'inconvénient des deux premiers formats est que l'outil *Microsoft Excel* de la suite bureautique *Microsoft Office* peut confondre les symboles. Par exemple, en anglais, la séparation des colonnes se fait avec des virgules et les décimales des nombres sont représentées par des points tandis qu'en français, les colonnes sont réalisées avec des points virgules et les décimales sont représentées par des points.

I.3.2/ Phase de post-production

À la jonction du tournage et la post-production, il existe des utilitaires permettant de modifier les métadonnées. Ces logiciels permettent de les éditer, de générer de nouveaux rapports son, de permuter les pistes entre elles, de créer des fichiers polyphoniques et monophoniques ou encore d'extraire uniquement les sons seuls. Voici une liste non exhaustive des logiciels disponibles actuellement sur le marché :

- *Wave Agent* de Sound Devices ;
- *Majax d'Aaton* : destiné spécifiquement aux rushes générés par le Cantar-X, il permet en plus des fonctions présentées ci-dessus de modifier le mixdown en reprenant les niveaux de chacune des pistes ;
- *Polynamer* développé par Polyson ;
- *BWF MetaEdit* développé par la Federal Agencies Digitization Guidelines Initiative (FADGI) ;
- *BWF Manager* de Fostex ;
- *Bwf-Widget* de GoodSound ;
- *Metadata Touch* de Digital Confidence.

Au moment du montage son, on retrouve d'autres logiciels d'édition ou de gestion des *rushes*. Toutefois, la finalité reste la même dans chaque processus de montage : il s'agit de trouver un fichier son désiré le plus simplement et rapidement possible. Par exemple, si sur l'une des prises, l'acteur vient à rater sa réplique, le monteur paroles³² peut rechercher un double dans les *rushes* en reprenant

³² Le monteur-parole est un métier de la chaîne du son en postproduction. Intervenant à la suite du montage image, il fait le lien entre la prise de son et le mixage, en rendant le plus exploitable possible le son issu du tournage.

une prise dont les numéros de séquence et de plan sont identiques à la prise en question. Nous présentons ci-dessous, de manière succincte et non exhaustive, les logiciels les plus employés :

- *Soundminer et BaseHead ;*

Utilisés en parallèle d'une STAN, ces deux logiciels intègrent différentes tranches comme le iXML. Appelés moteurs de recherche, ils permettent d'accéder rapidement à des fichiers audio. L'avantage de ces logiciels est d'offrir un environnement personnalisable et simple d'utilisation. Les recherches reposent soit sur la requête de termes dans les fichiers audio mêmes soit sur l'usage de bases de données³³ internes au logiciel. Enfin,

- *Adobe Audition, Pyramix, Nuendo et Pro Tools.*

Pro Tools offre à l'utilisateur un espace dédié à l'écoute et au tri des *rushes*, le *Workspace*. Il reprend les deux méthodes de recherche citées ci-dessus. À noter qu'il utilise non pas les tranches BEXT ou iXML mais la tranche ID3v2, conçue au départ pour les formats MPEG, pour renseigner les métadonnées. Cependant, le choix de cette tranche peut engendrer des problèmes de compatibilité entre les différentes plateformes de travail.

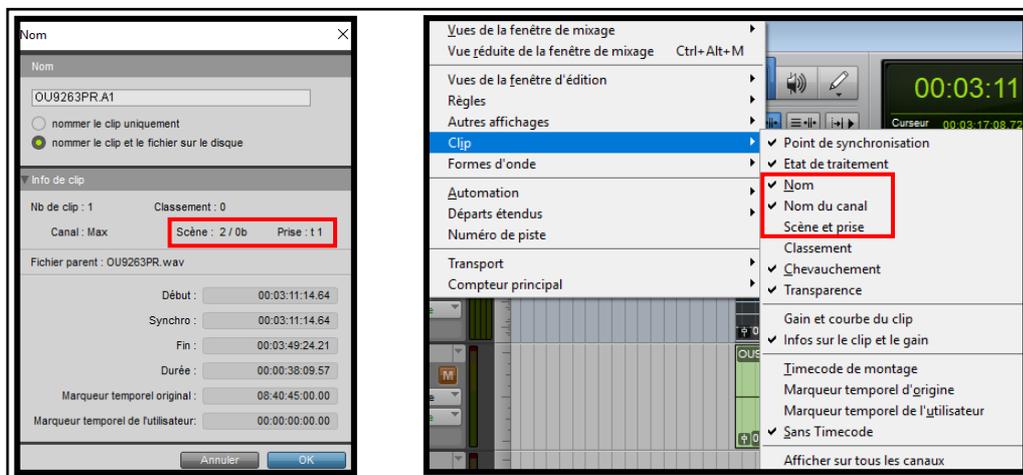


Figure 1.10 - Captures d'écran de l'affichage des métadonnées dans Pro Tools (Version 12.7.1) (à gauche les informations sont disponibles pour un clip spécifique ; à droite le menu permet d'afficher différentes métadonnées de montage image et de tournage)

³³ Une *base de données* correspond à un assemblage de données extraites de fichiers et stockées dans un fichier unique et strictement organisé. Cette méthode permet de rechercher une ressource rapidement et d'y accéder ultérieurement. Le fait d'employer des bases de données permet d'éviter de scanner l'ensemble des fichiers à chaque démarrage du logiciel.

Les métadonnées considérées comme très importantes pour la post-production apparaissent en premier dans l'affichage dans Pro Tools (*Figure 1.10*). En renseignant ces quelques métadonnées, le monteur paroles et le monteur son ont à disposition toutes les informations indispensables à la bonne compréhension du dispositif de prise de son.

Dans ce premier chapitre, nous avons analysé le développement des métadonnées professionnelles en audio au tournant des années 2000 et exploré leur potentiel concret dans la gestion des *rushes* provenant de l'étape du tournage. Il s'avère que la tranche iXML est désormais très bien implantée dans le secteur de l'audiovisuel. Son utilisation permet de simplifier des processus extrêmement complexes comme la phase de conformation. L'avènement et l'essor de la tranche iXML demande à reconsidérer l'emploi du rapport son manuscrit, ce qui pourrait accélérer son développement. Toutefois, le manque de normalisation et la grande disparité en termes d'implémentation chez les différents fabricants pourraient fragiliser à long terme la tranche iXML.

Toutes les métadonnées de la tranche iXML ne correspondent pratiquement qu'à des métadonnées statiques. Aucune métadonnée actuelle n'informe sur les paramètres de traitement audionumérique ayant lieu avant ou après la Conversion Analogique-Numérique (CAN). Aucune information sur les mouvements de *faders*³⁴, le filtrage ou la phase, qui sont pourtant des données irréversibles et inscrites de manière indélébile dans le signal, ne sont renseignées dans le domaine public du iXML.

Afin de mieux comprendre les différents enjeux qui découlent de la non-exploitation de ces informations, un état des lieux de la prise de son en France est proposé dans le *Chapitre II*. À travers ces analyses, nous développerons et justifierons l'utilisation de nouvelles métadonnées.

³⁴ Le terme anglais *fader* désigne en électronique les potentiomètres rectilignes à glissière.

Chapitre II

Pourquoi exploiter de nouvelles métadonnées

L'implémentation de nouvelles métadonnées doit répondre et pallier à de réelles problématiques de terrain. L'objectif de cette approche est de simplifier la jonction entre le monde de la production et de la post-production. Parallèlement, générer de nouvelles informations c'est aussi prendre le risque de complexifier les méthodes de travail.

Cette seconde partie a été réalisée à l'aide d'entretiens (*cf.* questionnaires en *Annexe F*) auprès de professionnels représentant les différents corps de métier de la chaîne du son au cinéma : chefs opérateurs du son, monteurs parole, monteurs son et mixeurs. Nous proposons d'analyser dans un premier temps les éléments sonores impliqués dans l'emploi de nouvelles métadonnées. Dans un deuxième temps, nous présenterons un historique des enregistreurs et des problématiques liées au *mixdown*. Enfin, dans un troisième temps, nous présenterons une vision globale des pratiques actuelles de la prise de son en France en analysant les différents traitements utilisés lors de la prise de son.

II.1/ Le son direct

En France, une grande importance est attachée au « son direct ». Ce terme renvoie à « *tous les enregistrements qui sont effectués durant la phase de tournage. Au sens strict du terme, il s'agit uniquement des sons synchrones à l'image, enregistrés pendant la prise de vue correspondante. En pratique, qu'ils soient associés à une prise de vue ou captés séparément, on les classe dans cette famille pour les différencier des sons additionnels qui seront ensuite fabriqués durant le montage, le bruitage ou le mixage.* »³⁵. À travers cette définition, il est rapporté que le son direct regroupe majoritairement les dialogues du film et diverses familles de sons : ambiances enregistrées sur les lieux du tournage, sons d'accessoires et musiques enregistrées en direct. Les dialogues sont d'une extrême importance car c'est à travers eux que les émotions des comédiens et le sens de l'histoire sont véhiculés. Dans cette partie, nous analyserons la place de la voix dans le cinéma français et la notion d'intelligibilité.

³⁵ Lucien Balibar, *La chaîne du son au cinéma et à la télévision : De la prise de son à la post-production*, *op. cit.*, p. 3.

II.1.1/ Importance de la voix dans la bande son

La voix au cinéma est « apparue » en 1927, date convenue de l'apparition du cinéma parlant, à travers le film *The Jazz Singer* d'Alan Crosland, produit par la société Warner Bros. Depuis cette date historique, la voix n'a cessé d'être employée à l'écran.

La voix est en premier lieu représentative d'un milieu socio-culturel. À travers le vocabulaire, les tournures de phrases, les expressions régionales qui la composent, elle est un marqueur social. La voix est un objet complexe qui ne se compose pas uniquement de dialogues, on parle de « *geste vocal* »³⁶. Entre le souffle, le cri ou le chuchotement du comédien, la voix permet d'accéder à un « *monde moteur miniature, [...] un état intérieur* »³⁷. Elle « *communique une pensée, une intention, cherche à influencer. La manière de parler est souvent plus importante que les choses dites.* »³⁸. Chaque élément émis par la bouche d'un comédien peut être source de fonction narrative voire de stratégie communicative. De plus, le geste vocal énonce parfois une réalité différente de celle proposée à l'image. Le spectateur peut alors percevoir des émotions et des infimes clés de compréhension. Ces multiples facettes rendent la voix déterminante dans le contexte cinématographique.

*« Tout le monde sait intuitivement que la voix est aussi caractéristique que le visage, que nous nous sommes tous retournés un jour pour reconnaître « de visu » un timbre que nous avons identifié à l'oreille [...] La voix humaine est la source principale d'un film, c'est aussi la plus variée. »*³⁹ (Lucien Balibar)

L'hégémonie de la voix sur les autres sons du film n'est pas seulement due à un problème d'intelligibilité. La voix hiérarchise les sons autour d'elle. L'effet naturel de « vococentrisme »⁴⁰ lié à l'écoute binaurale humaine est devenu aussi l'écoute du cinéma parlant. Michel Chion propose une analyse personnelle de la voix au sein de la bande sonore filmique. Selon lui, « *il y a les voix, et tout le reste. [...] dans n'importe quel magma sonore, la présence d'une voix humaine hiérarchise la perception autour d'elle* »⁴¹. Pour argumenter la notion de « vococentrisme », il prend pour exemple le réflexe naturel de favoriser la voix et le texte qu'elle porte au cinéma. On renforce artificiellement le niveau

³⁶ Claude Baiblé (sous la dir. de), *Le son au cinéma*, Paris, Ministère de l'Éducation et de la Culture, coll. Langage et techniques, 1993, p. 111.

³⁷ Claude Baiblé (sous la dir. de), *ibid*, p. 111.

³⁸ Claude Baiblé (sous la dir. de), *ibid*, p. 111.

³⁹ Lucien Balibar, *La chaîne du son au cinéma et à la télévision : De la prise de son à la post-production*, op. cit., p. 8.

⁴⁰ Michel Chion, *La Voix au Cinéma*, Paris, Cahiers du Cinéma, coll. Essais, 1982, p. 19.

⁴¹ Michel Chion, *ibid*, p. 19.

et la présence de la voix par rapport aux autres sons afin de reproduire cet isolement naturel vis-à-vis du bruit environnant.

« Ce que l'on appelle, sur le tournage, la prise de son est en réalité dans la majorité des cas une « prise de voix », les autres bruits étant dans la mesure du possible adoucis ou éliminés. »⁴².

En France, la qualité du son direct est très souvent reconnue à l'étranger. On pourrait dater l'attachement des chefs opérateurs du son au son direct aux années 1960, au moment de la Nouvelle Vague où « les chefs opérateurs du son deviennent des acteurs centraux dans l'apport de matière sonore au film »⁴³. Courant cinématographique français, la Nouvelle Vague fait date dans l'histoire du cinéma hexagonal par la production d'un grand nombre de films tournés en extérieur. Ces nouvelles méthodes de travail ont permis d'apporter une nouvelle esthétique, plus proche du réel. Ce courant est aussi lié à l'évolution des technologies avec la conception de caméras légères et silencieuses comme la caméra *Eclair 16* par l'ingénieur français André Coutant de la société Éclair et l'apparition du magnétophone portable à bandes Nagra construit par Stefan Kudelski remplaçant les encombrants camions son de tournage. C'est au cours de cette époque que le *mythe du son direct* s'est installé. Néanmoins, les films de la Nouvelle Vague ne se basent pas tous sur le son direct. Par exemple, les films majeurs *A bout de souffle* de Jean-Luc Godard ou encore *Paris nous appartient* de Jacques Rivette et *Zazie dans le métro* de Louis Malle ont été entièrement postsynchronisés. Cependant, cette époque a permis aux chefs opérateurs de développer une grande exigence vis-à-vis du son direct. Désormais, les « sons sonnent le réel ! »⁴⁴.

II.1.2/ Problématique de l'intelligibilité au cinéma

La question de l'intelligibilité est au cœur de débats en France. Encore de nos jours, elle reste une vraie problématique. Il ne s'agit pas de la qualité de la bande son mais de sa compréhension. « Qu'est-ce qu'il a dit ? ». Cette phrase nous ramène au spectateur regardant un film au cinéma ou à la télévision en famille et ne comprenant pas certaines répliques du film. Pourtant, aux regards des technologies actuelles, il semble difficile de trouver la source du problème. La qualité technique de la

⁴² Michel Chion, *La Voix au Cinéma*, Paris, Cahiers du Cinéma, coll. Essais, 1982, p. 19.

⁴³ Olivier Guillaume, *L'organisation en postproduction sonore cinématographique*, mémoire sous la direction de Jean-Pierre HALBWACHS et Selim AZZAZI, École Nationale Supérieure Louis Lumière, section son, 2010, p. 61.

⁴⁴ Christian Canonville, « De quelques implications esthétiques des innovations techniques dans la dimension sonore des films », *Cahier Louis-Lumière* n°1, automne 2003, p. 94.

captation de la voix est intimement liée aux technologies et donc aux époques : elle n’a jamais cessé d’augmenter.

Depuis les années 1950 et la généralisation de l’enregistrement sur bandes magnétiques, la qualité de la voix n’a cessé d’augmenter. Désormais, l’enregistrement sur disque dur, où la quantification est fixée à 24 bits, propose une dynamique de l’ordre de 144 dB !

Néanmoins, malgré les évolutions technologiques très importantes, le reproche de l’intelligibilité est encore d’actualité. Par exemple, dans un article de Télérama intitulé « *Il y a des sons qui se perdent...* »⁴⁵, il est écrit :

*« Les films tiennent parfois de la bouillie sonore. La faute aux acteurs...
et au culte du son direct, hérité de la Nouvelle Vague. »*

À cette phrase choc, le mixeur français François Groult répond : « *depuis que je fais ce métier, j’entends ce reproche régulièrement* »⁴⁶. Alors qui est le responsable de ce manque d’intelligibilité des comédiens ? « *Les premiers responsables ? Ce sont les comédiens* »⁴⁷ répond Gérard Lamps. Selon lui, « *il y a quarante ans, les comédiens articulaient, y compris pendant les passages chuchotés ou dans les scènes d’émotion. Et on comprenait tout...* »⁴⁸.

Ce constat a aussi été évoqué dans la presse quelques années plus tard. En 2015, à la fin du festival de Cannes, un article de Charlie Hebdo et plus récemment, un autre article du Canard Enchaîné de 2016 évoquaient la difficulté de compréhension des dialogues dans certains films, dont des films français. En Juin 2017, au Cinéma Les Fauvettes à Paris, un débat⁴⁹ sur l’intelligibilité, mené par la Commission Supérieure Technique de l’image et du son (CST) des dialogues, a été réalisé. En réaction à cette réunion de la CST, un autre débat⁵⁰ a été organisé par la société A4audio où le chef opérateur du son Stéphane Bucher et le mixeur Steven Ghouti exposent leurs points de vue. On retrouve les mêmes pistes de réflexion au sein de l’AFSI à travers l’atelier « Le son direct »⁵¹ réalisé en 2013.

⁴⁵ Aurélien Ferenczi, « Il y a des sons qui se perdent... », Télérama, décembre 2009, URL : <http://www.telerama.fr/cinema/il-y-a-des-sons-qui-se-perdent,50545.php>

⁴⁶ Aurélien Ferenczi, *ibid.*

⁴⁷ Aurélien Ferenczi, *ibid.*

⁴⁸ Aurélien Ferenczi, *ibid.*

⁴⁹ Consultable à l’adresse URL suivante : <https://www.cst.fr/la-cst/les-rencontres-de-la-cst/>

⁵⁰ Consultable à l’adresse URL suivante : https://www.youtube.com/watch?v=IhSvFRrPS_k.

⁵¹ Thierry Lebon, « Compte-rendu - L’atelier « Le son direct » », Association Française du Son à l’Image, juin 2013, URL : <https://www.afsi.eu/articles/15233-atelier-le-son-direct>

Au regard des réponses apportées au cours de l'ensemble de ces ateliers et réunions, l'attachement au son direct en France est certain. De par cette spécificité française, les différents acteurs du son proposent les pistes de réponses suivantes à la question sur l'intelligibilité :

- *Une spécificité de la langue française ;*

La particularité « tonale » de la langue française fait que la majorité des fins de mot ou de phrase baisse de niveau sonore alors que c'est dans cette dernière syllabe qu'est véhiculée l'émotion. Le français est une langue plate ne possédant pas d'accent tonique.

- *L'héritage du passé ;*

Le cinéma français se distingue par son attachement historique au son direct. Cette particularité française se retranscrit même dans les mixages. Par exemple, pour une séquence où les comédiens se situent au milieu d'une foule, François Groult dit : « *le cinéaste américain coupe le son d'ambiance, pour se concentrer sur les deux acteurs et leurs dialogues. Le cinéaste français, lui, ne s'y résout pas, au nom d'une certaine idée du réalisme. Au risque de rendre inaudibles les paroles. Pourtant, dans la vraie vie, au cœur d'un lieu bruyant, l'oreille sait sélectionner ce qu'elle a envie d'entendre. Au cinéma, le réalisme passe parfois par ce type de conventions.* »⁵².

C'est aussi la difficulté d'utiliser la post-synchronisation en France, procédé technique très largement employé dans les pays anglo-saxons. Comme le souligne Gérard Lamps : « *le son direct est un véritable culte, hérité des années 1960. Il est le garant du réalisme, mais celui-ci doit avoir des limites, ce que le cinéma américain a bien compris.* »⁵³.

- *Évolution des pratiques de tournage ;*

Le « pouvoir » du chef opérateur du son sur un plateau est en recul. Autrefois, il pouvait refuser un décor impropre à une bonne prise de son. Désormais, il devient même difficile de couper une prise. De plus, la participation aux repérages des chefs opérateurs du son est de moins en moins fréquente.

L'évolution des outils de post-production a aussi changé les mentalités. Désormais, on peut généralement convenir qu'une scène difficile sera rattrapable en post-production son avec notamment des outils de restauration sonore comme la *suite RX6* de la société Izotope.

⁵² Aurélien Ferenczi, « Il y a des sons qui se perdent... », *op. cit.*

⁵³ Aurélien Ferenczi, *ibid.*

- *Le comédien.*

Le constat est le suivant : les comédiens parlent de moins en moins fort. « *Le niveau de leur voix est inversement proportionnel à l'évolution technique des appareils enregistreurs !* »⁵⁴. Les causes sont multiples. Le phénomène est lié à l'avènement des microphones cravates sur les plateaux de tournage. Sachant que ce type de microphone capte au plus près le signal de la voix, les comédiens ont tendance à parler moins fort. Les situations techniques peuvent aussi les mettre en difficulté comme les prises en extérieur qui demandent un « porter de voix » supérieur à une séquence filmée en plateau dans les studios. Au sujet des microphones cravates, on retrouve ces considérations dans les propos de différents chefs opérateur du son :

« Dans les années cinquante, les acteurs jouaient souvent de façon très théâtrale. Même en claquant une porte, ils déclamaient – ce qui n'est absolument pas logique – en disant : « Zut, j'ai oublié mes clefs ! ». Aujourd'hui, on évolue vers un plus grand naturel : les contraintes techniques sont moins grandes, nous utilisons des micros plus sensibles, nous limitons moins les acteurs dans leur façon de jouer... Cela fait partie de l'intérêt du micro HF⁵⁵. Je détesterais aller voir un acteur pour lui demander de parler plus fort, j'aurais l'impression de ne pas faire mon boulot. »⁵⁶
(Éric Devulder)

« Pour la Reine Margot, je suis allé plusieurs fois dire des choses directement à Dominique Blanc ou de temps en temps à Isabelle Adjani et Vincent Perez que je connaissais pour avoir déjà tourné plusieurs fois avec eux. Je me suis retrouvé en porte-à-faux avec Patrice Chéreau à propos d'une scène entre Isabelle et Dominique entourées de trois cent cinquante figurants. Dominique jouait cette scène de façon très limite, très fragile. Je suis allé la voir après la première prise et lui ai rappelé que la figuration faisait semblant de parler, il y aurait de la musique, ce serait une fête bruyante... et que j'avais énormément de difficulté à l'entendre donc... Elle m'a répondu : « Patrice vient justement de me dire qu'il fallait que je parle un peu plus bas » »⁵⁷ (Guillaume Sciamma)

⁵⁴ **Thierry Lebon**, « Compte-rendu - L'atelier « Le son direct » », Association Française du Son à l'Image, juin 2013, URL : <https://www.afsi.eu/articles/15233-atelier-le-son-direct>

⁵⁵ Le terme « micro HF » est très généralement employé pour désigner les microphones cravates sans fil dont la liaison est réalisée par transmission hertzienne (Hautes fréquences).

⁵⁶ **Claudine Nougaret, Sophie Chiabaut**, *Le son direct au cinéma*, Paris, FEMIS – Institut de Formation et d'enseignement pour les métiers de l'image et du son, 1997, p. 92.

⁵⁷ **Claudine Nougaret, Sophie Chiabaut**, *Le son direct au cinéma*, op. cit., p. 94.

« L'invention de la liaison microphonique HF, pensée initialement pour quelques idées spécifiques de mise en scène, a permis d'obtenir une intelligibilité identique quelle que soit la valeur du plan [...]. De fil en aiguille, plus aucun acteur de la jeune génération ne parle suffisamment fort pour être audible sur le plateau, et tout juste assez pour le micro qu'il porte. Il arrive fréquemment que même le perchman ou le cadreur, qui sont pourtant bien placés au moment de la scène, n'entendent pas ce que l'acteur dit ! Et désormais il est bien délicat de demander à un acteur de forcer un peu sur son organe sous peine de passer pour un grossier technicien incapable de respecter la sensibilité du jeu. »⁵⁸ (Lucien Balibar)

L'ensemble de ces témoignages renvoie à l'exigence du son direct en France et l'amour porté aux voix des comédiens. Cette voix, « vecteur de la compréhension du film »⁵⁹, occupe une place singulière et est au centre des exigences du chef opérateur du son. Pierre Gamet, interrogé sur quel serait son film idéal, répondait : « J'ai un rêve très simple : qu'un metteur en scène donne de l'importance à d'autres sons que la parole, qu'il ait une autre réflexion, une autre approche de la bande son et qu'il s'attache à un bruit, à un silence »⁶⁰.

II.2/ Avènement de l'enregistreur multipiste

Au moment de l'enregistrement du son direct, le chef opérateur du son en est le responsable. Aidé de son assistant, nommé dans la pratique *perchman*, il enregistre les différentes sources captées par les microphones placés sur le plateau du tournage. Les méthodes pour enregistrer ces différentes sources sont diverses et variées et résultent de méthodes de travail élaborées au fil des années. Dans cette partie, nous analyserons l'évolution des enregistreurs et des problématiques qui en découlent.

⁵⁸ Lucien Balibar, *La chaîne du son au cinéma et à la télévision : De la prise de son à la post-production*, op. cit., p. 12.

⁵⁹ François Groult, « Compte rendu - Traitement des dialogues en auditorium », Association Française du Son à l'Image, décembre 2008, URL : <https://www.afsi.eu/articles/22581-compte-rendu-traitement-des-dialogues-en-auditorium>

⁶⁰ Claudine Nougaret, Sophie Chiabaut, *ibid*, p. 189.

II.2.1/ Historique des enregistreurs

Dans des situations « nomades », le matériel de prise de son doit être robuste, mobile et flexible aux différentes conditions rencontrées. Les outils d'enregistrement évoluent et modifient les pratiques mises en place. Les méthodes de travail sont constamment remises en question. Au prix d'une complexité de plus en plus lourde, l'enregistreur actuel se propose d'être un outil fabriquant la matière la plus transversale qui soit.

L'enregistreur audionumérique sur disque dur est l'aboutissement de longues années d'évolution de l'enregistrement du son. En 1927, lors de la sortie de *The Jazz Singer* d'Alan Crosland, le procédé *Vitaphone* est alors employé. Ce procédé d'enregistrement et de diffusion synchrone du son développé par la compagnie Western Electric en 1926 est alors l'innovation la plus aboutie dans le domaine du son depuis une dizaine d'années. Toutefois, à la fin des années 1920, c'est l'enregistrement optique sur pellicule du son sur film photographique qui s'impose. Ce procédé utilise alors comme principe de diffusion le *Fox-Movietone* alors concurrent du *Vitaphone*. Au même moment, « toutes les compagnies se mettent à produire du cinéma parlant, fin 1929 »⁶¹.

L'autre grande révolution technique dans le domaine du son est l'enregistrement sur bandes magnétiques. En 1948, les recherches réalisées durant la guerre permettent de mettre au point cette technologie aux Etats-Unis avec la RCA (*Radio Corporation of America*) et en Allemagne avec l'AEG (*Allgemeine Elektrizitäts-Gesellschaft*). Le son optique n'est plus employé qu'en copies d'exploitation pour les salles de cinéma. Le son sur bandes magnétiques présente les avantages suivants :

*« La diminution de la distorsion, l'accroissement de la dynamique des sons, l'élargissement de la bande passante, la lecture immédiate du résultat de l'enregistrement, le bruit de collure qui devenait inaudible en utilisant une coupe en biais du film »*⁶².

Au niveau des fabricants, la société Nagra propose dès 1953, un enregistreur portable sur bandes magnétiques, le Nagra II. Fonctionnant sur batterie, le chef opérateur du son gagne en autonomie et « la synchronisation avec l'image est assurée par un signal pilote (provenant de la caméra) enregistré

⁶¹ **Martin Barnier**, « Les premiers ingénieurs du son français », 1895. *Mille huit cent quatre-vingt-quinze*, Histoire des métiers du cinéma en France avant 1945, n°65, 2011, p. 5.

⁶² **LEROUGE, Claude**, *Sur les 100 années, le cinéma sonore*, Paris, Dujarric, 1996, p.134.

sur une piste prévue à cet effet (en plus du clap bien entendu) »⁶³. À la fin des années 1960, on assiste à une généralisation de l'usage du son magnétique.

Durant cette période, de nouvelles innovations voient le jour comme l'apparition des premiers enregistreurs portables stéréophoniques avec comme figure de proue le célèbre Nagra IV-S en 1971. Il permet d'enregistrer deux signaux séparément : on emploie le terme d'« enregistreur bipiste ». Cette innovation offre la possibilité de réaliser de nouvelles prises de son au « couple » comme pour les ambiances. Au cinéma, cette nouvelle technologie est très vite détournée et bouleverse les méthodes de travail. Au lieu d'enregistrer un couple stéréophonique sur les deux pistes, le chef opérateur du son réalise une séparation entre les signaux en fonction de leur degré d'importance. Par exemple, il place la perche, signal fondamental, sur la piste n°1 tandis que les autres signaux, comme les microphones cravates sont placés sur la piste n°2. Cette piste est ainsi un mélange de l'ensemble des signaux dit « moins nobles » provenant des microphones cravates. Ce détournement permet de protéger le signal de la perche des signaux provenant des microphones cravates.

Toutefois, cette nouvelle méthode de travail reprend les principes de l'enregistrement monophonique : le chef opérateur du son doit moduler les signaux en temps réel pour ne pas les bruyter entre eux. Par exemple, imaginons un plan avec deux comédiens A et B. Les signaux provenant du plateau sont les suivants : une perche principale et deux microphones cravates équipant chaque comédien. À chaque réplique du comédien A, le chef opérateur du son doit « enlever » le signal provenant du microphone cravate du comédien B et inversement.

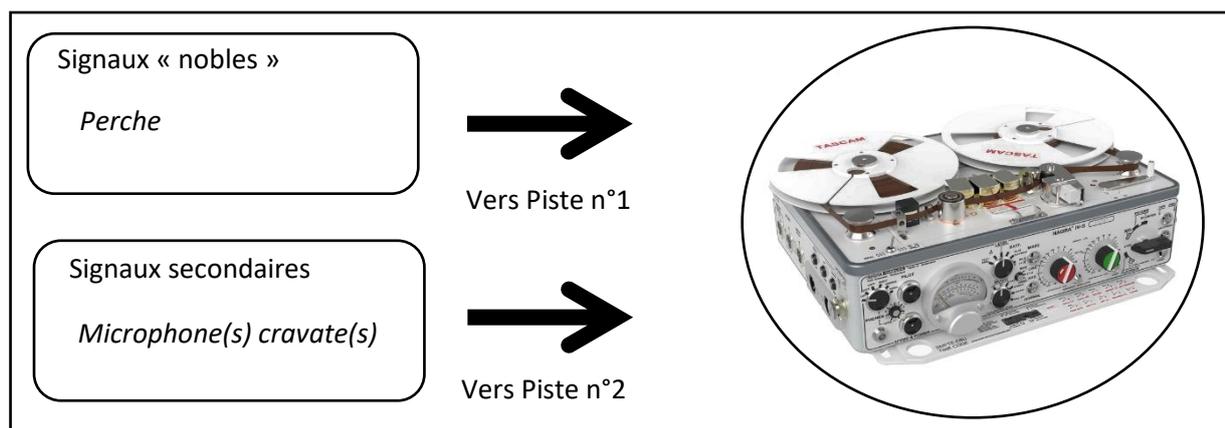


Figure 2.1 - Détournement possible de l'enregistreur Nagra IV-S

⁶³ Romain Vuillet, *De quelques nouvelles potentialités en postproduction cinématographique apportées par l'usage de l'enregistreur multipiste*, mémoire sous la direction de Daniel SOBRINO et Christian CANONVILLE, École Nationale Supérieure Louis Lumière, section son, 2006, p. 8.

De plus, la possibilité d'enregistrer deux signaux séparément permet de les dissocier temporellement. Par exemple, prenons le cas où nous enregistrons une personne avec une perche et un microphone cravate. La perche située à une distance variable du comédien en fonction de ses mouvements est enregistrée en même temps que le microphone cravate placé à une distance pratiquement fixe. Captant le même signal sonore mais placés à des distances différentes, ces deux signaux auront un décalage temporel. Or, lors du montage et du mixage, ces deux signaux seront écoutés en même temps ce qui engendrera un filtrage en peigne⁶⁴ inévitable. Du fait de pouvoir les isoler, on peut modifier leurs positions temporelles individuellement afin de compenser ce décalage au moment de l'enregistrement. C'est la « remise en phase ».

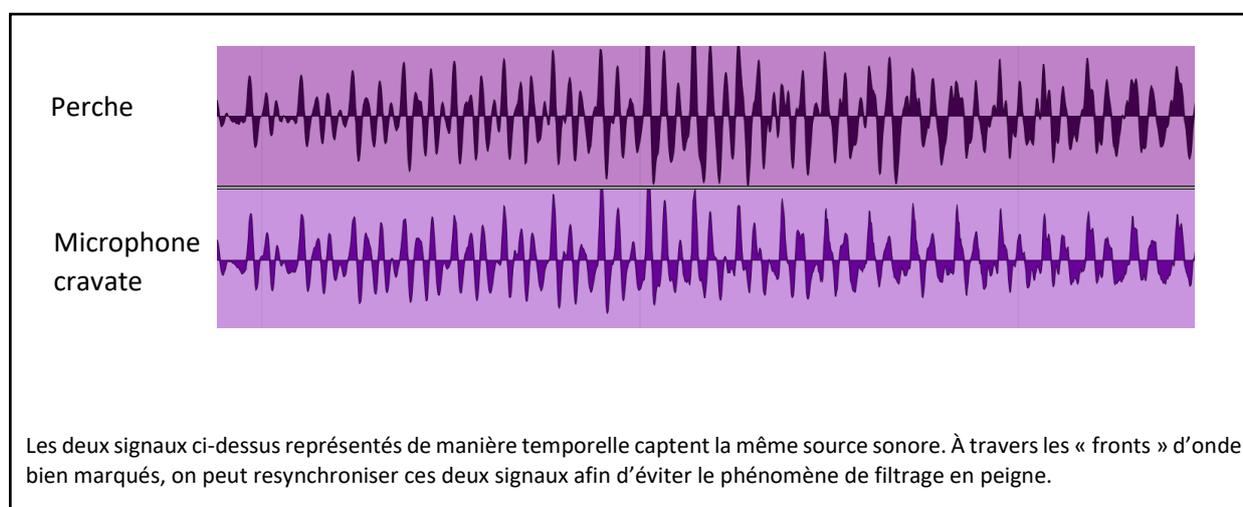


Figure 2.2 - Présentation de la méthode de remise en phase

Dès le milieu des années 1980, les premiers enregistreurs numériques sur bandes magnétiques voient le jour. Ils enregistrent alors sur deux voire quatre pistes.

« Le quatre pistes dans les années 1990 a provoqué un changement radical en permettant de séparer régulièrement les sources pour permettre leur traitement individuel en post-production. Il est le début d'un lien indissociable entre la prise de son et le montage paroles, qui a permis un bond en avant de la qualité technique de la bande sonore. »⁶⁵

⁶⁴ Le terme *filtrage en peigne* est employé en traitement du signal. Il correspond au mélange du signal original avec sa version retardée. Cette différence de temps crée des interférences destructives et constructives à intervalle de temps régulier. En pratique, ce filtrage colore le timbre d'une voix et se fait surtout sentir dans les basses fréquences.

⁶⁵ **Vincent Magnier**, « Compte-rendu « Exploiter le son direct du plateau à l'auditorium avec le nouveau Aaton Cantar X3 », Association Française du Son à l'Image, février 2016, URL : <https://www.afsi.eu/articles/24659-le-nouveau-aaton-cantar-x3>

Cette époque est représentative d'une non-homogénéité des formats audio. Du support numérique d'enregistrement au support de diffusion, les formats naissent et disparaissent au gré des innovations. Cette période correspond surtout à l'emploi de supports numériques d'enregistrement non-professionnels. Par exemple, le Nagra-D apparu en 1992 utilise comme support d'enregistrement le *Digital Audio Tape (DAT)*.

Comme une évidence, juste avant les années 2000, les enregistreurs deviennent l'équivalent d'« ordinateurs portables ». Cette dernière innovation, depuis totalement implémentée dans tous les enregistreurs du marché, a été portée par la société Zaxcom. En 1996, elle propose le tout premier enregistreur portable sur disque dur avec quatre pistes, le DEVA. Au cours des années 2000, les enregistreurs viennent à proposer jusqu'à 8 pistes enregistrables simultanément.

Le Cantar-X, à travers l'emploi d'un disque de 127 Go, permet d'enregistrer jusqu'à 33 heures de son en 48 kHz, 24 bits sur 8 pistes. C'est l'équivalent d'environ 200 heures de bandes DAT...

Désormais, du Cantar-X3 d'Aaton sorti en 2014 au DEVA 24 de Zaxcom sorti en début 2018, les constructeurs proposent 24 pistes. Difficilement toutes utilisables au même moment en condition de travail réel, l'enregistreur audionumérique de tournage semble avoir atteint ses limites en termes de nombre de pistes.

On peut observer que l'évolution des enregistreurs audio s'est jouée sur plusieurs niveaux :

- évolution du support d'enregistrement ;
- augmentation et diversification de la nature des signaux ;
Les enregistreurs actuels peuvent enregistrer différents types de signaux : analogique et numérique comme l'AES-42⁶⁶ ou encore le DANTE⁶⁷.
- augmentation considérable du nombre de sources enregistrables séparément.

II.2.2/ Pertinence et enjeu du *mixdown*

À l'ère du tout-numérique et de l'évolution des enregistreurs audio, des solutions pratiques de *workflow* ont été mises en place. Au moment du montage image, il serait absurde d'utiliser à chaque plan l'ensemble des pistes fournies par le chef opérateur du son. Afin de simplifier la gestion des *rushes* son de tournage, « *phase durant laquelle on s'occupe essentiellement de structure narrative et pas*

⁶⁶ L'AES-42 renvoie à une spécification officielle de l'AES datant de 2001. Cette norme régit le fonctionnement de la technologie des microphones numériques.

⁶⁷ Le terme DANTE correspond à *Digital Audio Network Through Ethernet*. C'est un assemblage de technologies normées et non un protocole. Il transmet des signaux audionumériques à travers le protocole Ethernet.

encore des détails techniques »⁶⁸, une solution a été très largement retenue : les enregistreurs audionumériques ont proposé une réduction de l'ensemble des pistes en deux pistes, en quelque sorte une version mixée et donc simplifiée de l'enregistrement, le *mixdown*. Dans un certain sens, le *mixdown* permet donc de patienter dans de bonnes conditions jusqu'à l'étape du montage paroles. Ce mixage témoin est important car c'est lors de l'étape du montage image que le réalisateur « se met le film dans l'oreille ». On pourrait penser qu'un *mixdown* monophonique serait une solution pertinente. Toutefois les microphones cravates génèrent des bruits parasites ainsi le monteur image peut les enlever ou les diminuer dans un souci de confort d'écoute.

Ce *mixdown*, mélange-réduction en français, est comparable à la matière fournie par les enregistreurs bipistes au temps du Nagra IV-S. Cette méthode est même appliquée dans les premières années des enregistreurs multipistes.

Afin de réaliser ce mélange en direct durant la prise, les chefs opérateurs du son emploient deux grandes familles d'équipement :

- les consoles analogiques ;

Elles sont en quelque sorte une version miniaturisée des consoles de concert actuelles. On retrouve un étage de préamplification, d'égalisation, d'envois à travers les auxiliaires et de mixage avec les *faders*. C'est avec cet outil que les chefs opérateur du son fabriquent le *mixdown*.



*Figure 2.3 - Photographies des consoles analogiques de tournage
(à gauche CS 208 V2 de Cooper Sound ; au centre SX-S6 de Sonosax ; à droite SX-ST8D de Sonosax)*

⁶⁸ Lucien Balibar, *La chaîne du son au cinéma et à la télévision : De la prise de son à la post-production*, op. cit., p. 158.

- les surfaces de contrôle.

Elles sont numériques et propriétaires. Majoritairement reliées à l'enregistreur en *USB*⁶⁹, chaque enregistreur est indissociable d'une surface de contrôle proposée par le fabricant. En 2018, les fabricants proposent les produits suivants : *CL-12* chez Sound Devices, *Oasis* chez Zaxcom et *Cantarem-2* et *Cantaress* chez Aaton.



*Figure 2.4 - Photographies des surfaces de contrôle de tournage
(à gauche Mix-16 de Zaxcom ; au centre CL-12 de Sound Devices ; à droite Cantarem2 d'Aaton)*

D'après des chefs opérateurs du son, avec qui ont été réalisés les entretiens, la fabrication du *mixdown* se réalise de la manière présentée ci-dessous. Notons que les pistes concernant les couples stéréophoniques ou les prises de son multicanal ne sont pas présentes dans le *mixdown*.

- Piste n°1 ou Gauche ;

Composée de : Perche principale et microphones d'appoint⁷⁰.

Fonction : Elle correspond à l'écoute naturelle. Dissocié de tous les autres signaux, le signal de la perche est « protégé ».

⁶⁹ Le terme USB correspond à *Universal Serial Bus* (en français bus universel en série). C'est une norme correspondant à un bus informatique série utilisé pour connecter des périphériques informatiques à un ordinateur.

⁷⁰ Les microphones d'appoint sont des microphones cachés dans le décor du tournage. Ils permettent de couvrir des zones sonores difficilement atteignables pour le *perchman*. En fonction des situations, ils sont reliés à l'enregistreur par liaison filaire ou hautes-fréquences.

- Piste n°2 ou Droite.

Composée de : tous les microphones cravates, tous les autres signaux.

Fonction : Cette piste correspond à la « présence accentuée » des comédiens. Elle permet de réaliser des premiers choix au moment du montage image en jouant sur le niveau de ces signaux captant un signal de grande proximité. Cette piste regroupe aussi les signaux représentant une forme de risque au moment de l'enregistrement. Par exemple, les microphones cravates sont reliés à l'enregistreur par liaisons hautes fréquences (HF) et sont parfois inutilisables à cause de frottement contre les vêtements des comédiens. La technologie HF peut occasionner des décrochages (perte totale du signal) ou des perturbations importantes ainsi ces signaux sont séparés du signal principal de la perche. Selon les méthodes des chefs opérateur du son, les appoints peuvent être placés dans la piste n°2.

Au vu des exigences de la prise de son et de son importance dans la bande son finale, le chef opérateur du son devait alors réaliser un enregistrement extrêmement bien abouti dès la prise.

« Pour résumer tout ce qui n'est pas maîtrisé suffisamment trouve sa place dans la piste 2, alors que la piste 1 est la source de continuité indispensable à l'écoute naturelle en montage image »⁷¹ (Lucien Balibar)

Parallèlement à la création du *mixdown*, les chefs opérateurs affectent chaque source à une piste divergée⁷² (ISO) de l'enregistreur. Au cours de l'atelier « Exploiter le son direct du plateau à l'auditorium avec le nouveau Aaton Cantar-X3 »⁷³ de l'AFSI, un ordre prédéfini des sources a été donné afin d'éviter une trop grande variété de méthodes de travail avec les enregistreurs actuels. Voici l'ordre proposé :

- *mixdown* : Pistes 1 et 2 ;
- perche(s) : Piste 3 (pistes 4 et 5 avec l'emploi d'une deuxième perche) ;
- microphones d'appoint ;
- microphones cravates ;
- prise de son stéréophonique voire multicanal.

⁷¹ Lucien Balibar, *La chaîne du son au cinéma et à la télévision : De la prise de son à la post-production*, op. cit., p. 160.

⁷² Une piste divergée est une piste enregistrée ne correspondant pas au *mixdown*.

⁷³ Vincent Magnier, « Compte-rendu « Exploiter le son direct du plateau à l'auditorium avec le nouveau Aaton Cantar X3 » », Association Française du Son à l'Image, février 2016, URL : <https://www.afsi.eu/articles/24659-le-nouveau-aaton-cantar-x3>

Le *mixdown* des enregistreurs multipistes a été conçu comme « *une version témoin du son direct, une référence utile aux projections de rushes et au montage image* »⁷⁴. Or, de par les méthodes de travail de la prise de son en France, le *mixdown* était en quelque sorte déjà réalisé par les chefs opérateur du son dès l'arrivée sur le marché des enregistreurs bipistes. Au temps de la bande magnétique, le chef opérateur du son réalisait une sommation des signaux. Cette version monophonique ou bipiste fournie par le chef opérateur du son était pratiquement une version définitive. L'objectif était alors de fournir la matière la plus propre possible sachant qu'elle serait écoutée durant plusieurs semaines au cours de la post-production. Ce mélange, équivalent au *mixdown* actuel, représentait les choix et l'orientation esthétique souhaités par le chef opérateur du son en direct.

*« Je considère le son que je fais comme définitif. J'ai toujours appréhendé la prise de son de cette façon. Quand j'ai commencé, je n'avais que trois micros 160, je faisais des films où les productions n'avaient pas d'argent pour postsynchroniser. J'ai toujours mélangé et donné mon son comme je l'entendais avec mes oreilles, avec mon empreinte, mon grain. C'est un équilibre que j'ai choisi au moment où nous tournons cette prise-là et sur lequel effectivement on ne peut plus revenir. C'est dangereux mais j'en assume la responsabilité. [...] L'ingénieur du son qui mixera ne pourra rien changer à cet équilibre. En bipiste, il le pourrait, mais cela m'agacerait de savoir qu'il va trafiquer les HF. »*⁷⁵ (Pierre Gamet)

De par cette ressemblance entre le *mixdown* actuel et l'enregistrement bipiste au temps du Nagra IV S, au cours des années 2000, de nombreux débats ont écloso concernant les méthodes de travail. Différents professionnels ont échangé entre les années 2008 et 2012 sur le forum de l'AFSI autour de la question des pistes divergées et de la gestion du *mixdown* :

- « Montage des directs...Ou démontage du direct ? »⁷⁶ (2008) ;
- « Montage des Dialogues d'après Multipiste »⁷⁷ (2008) ;
- « Montage de directs et grosse fatigue »⁷⁸ (2010) ;
- « Mixdown du direct »⁷⁹ (2012).

⁷⁴ Lucien Balibar, *La chaîne du son au cinéma et à la télévision : De la prise de son à la post-production*, op. cit., p. 158.

⁷⁵ Claudine Nougaret, Sophie Chiabaut, *Le son direct au cinéma*, op. cit., p. 69-70.

⁷⁶ Consultable à l'adresse URL suivante :

<http://archive.afsi.eu/forum/montage-des-directsou-d%E2%88%9A%C2%A9montage-du-direct>

⁷⁷ Consultable à l'adresse URL suivante :

<http://archive.afsi.eu/forum/montage-des-dialogues-d-apr%E2%88%9A%C2%AEs-multipiste>

⁷⁸ Consultable à l'adresse URL suivante : <http://archive.afsi.eu/forum/montage-de-directs-et-grosse-fatigue>

⁷⁹ Consultable à l'adresse URL suivante :

<http://archive.afsi.eu/node/4223>

Comme pour les prises de son monophonique, les enregistreurs bipistes ont relativement peu changé le travail des voix. « *Le mixeur [pouvait] désormais un peu plus agir sur le direct puisqu'il disposait de signaux différenciés pour la perche et les microphones HF s'autorisant ainsi une balance différente de celle de l'ingénieur du son même si on peut noter qu'en général ces derniers tenaient à ce que l'on respecte les niveaux d'enregistrements respectifs des pistes qui représentaient alors leurs idée du rapport perche / HF.* »⁸⁰.

Désormais, dans le domaine du cinéma, il semble convenu que le *mixdown* sert de référent au moment de la projection des bobines et permet d'écouter les intentions du chef opérateur du son. Toutefois, les monteurs paroles et les monteurs son utilisent dans la majorité des cas les pistes divergées. Ne pouvant dissocier les sources entre elles, le *mixdown* empêche la remise en phase pour des configurations à plusieurs microphones cravates. Dans la pratique, en fonction des temps de travail et de la date fixée pour la journée de détection, le travail peut être réalisé à partir du *mixdown*.

Enfin, les conditions de tournage ont évolué. Actuellement, on retrouve souvent des configurations techniques avec plusieurs caméras où il y a peu de répétitions avant de tourner. Il est difficile de réaliser un *mixdown* propre et de manière optimale.

*« En tournage, où l'on tend souvent à vouloir couvrir tous les axes, tous les plans, toutes les durées, même en sachant que la plupart ne seront pas utilisés, l'opérateur est incité à aborder son travail de manière passive, au plan par plan et pas à l'échelle de la séquence. »*⁸¹

⁸⁰ **Romain Vuillet**, *De quelques nouvelles potentialités en postproduction cinématographique apportées par l'usage de l'enregistreur multipiste*, op. cit., p. 16.

⁸¹ **Vincent Magnier**, « Compte-rendu « Exploiter le son direct du plateau à l'auditorium avec le nouveau Aaton Cantar X3 » », Association Française du Son à l'Image, février 2016, URL : <https://www.afsi.eu/articles/24659-le-nouveau-aaton-cantar-x3>

II.3/ Traitements à la prise

À travers l'emploi d'une console analogique de tournage ou d'une surface de contrôle numérique, le chef opérateur du son dispose d'un certain nombre de traitements applicables au signal enregistré. Ces traitements sont décomposables en deux grandes familles : traitements fréquentiels et traitements dynamiques.

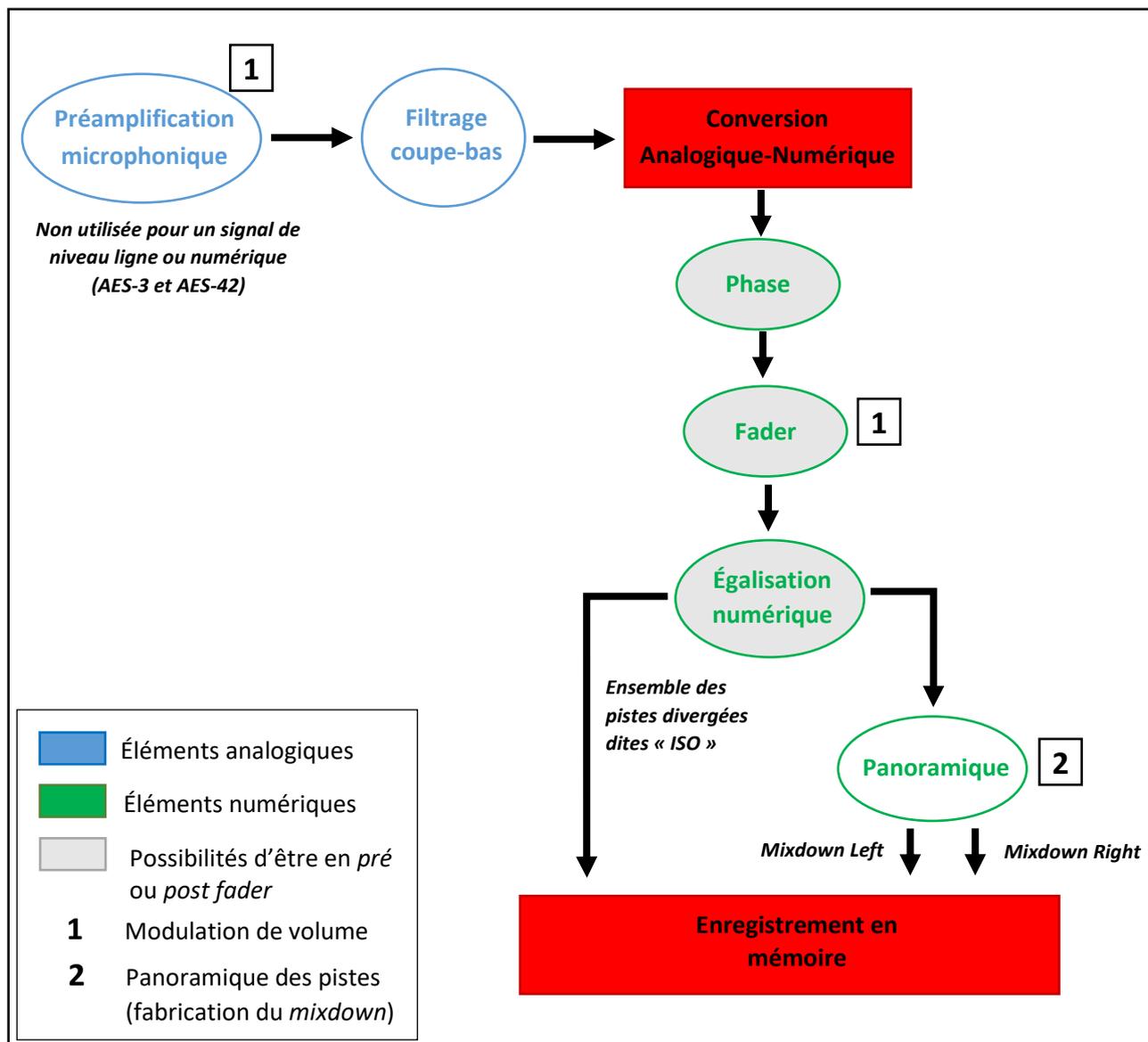


Figure 2.5 - Chaîne du son d'un signal microphonique avant et après la Conversion Analogique-Numérique (CAN) d'un enregistreur audionumérique

La Figure 2.5 ci-dessus reprend et synthétise les informations de l'Annexe G présentant la chaîne du son dans un Cantar-X3. On observe que certains traitements sont utilisés tant en analogique qu'en numérique dans un enregistreur audionumérique. Cette chaîne du son pratiquement identique dans

tous les enregistreurs actuels apporte une grande souplesse. Les méthodes présentées par la suite ne sont pas figées ; en fonction des conditions de tournage et du matériel à disposition, le chef opérateur du son peut les employer alternativement.

II.3.1/ Traitements fréquentiels : le filtrage

La famille des traitements fréquentiels se réfère au filtrage et aux égalisations réalisés au moment de l'enregistrement. Le fait d'être analogique ou numérique est exclusivement lié aux machines employées par le chef opérateur du son. Si ce dernier utilise une console analogique avant d' « attaquer » les entrées de l'enregistreur, alors ces traitements sont tous analogiques tandis que si le chef opérateur du son utilise une surface de contrôle, alors le filtrage est analogique et/ou numérique et l'égalisation numérique.

Le *filtrage analogique* est le premier traitement « obligatoire ». Exclusivement de type coupe-bas, il évite d'altérer le fonctionnement du préamplificateur microphonique. Par exemple, une énergie trop importante dans les fréquences situées dans l'extrême grave peut saturer l'entrée de l'enregistreur ou générer des aberrations dans le signal enregistré. Le filtrage de type coupe-bas est principalement employé pour atténuer les fonds sonores possédant des fréquences indésirables dans le grave. Aussi, il peut être utilisé dans la variation de l'effet de présence des comédiens.

*« En projection de rushes, je trouve qu'il y a trop de basses - c'est une phobie, j'ai du mal à couper des basses à la prise de son, j'en laisse toujours trop. J'aime bien les sons ronds, les sons qui ont du corps. Je préfère que le mixeur puisse tailler dedans après. »*⁸² (Pierre Gamet)

Lors de l'atelier « Le son direct » de l'AFSI, la question « *couper trop dans le bas lors de la prise de son ?* » a été soulevée. C'est lorsque le chef opérateur du son décide de couper dans les fréquences graves afin d'atténuer le fond sonore que l'outil employé n'est pas le bon. Des logiciels plus précis, comme ceux destinés à la restauration sonore sont à disposition des monteurs parole et des mixeurs et permettent de corriger le signal avec une plus grande précision, c'est-à-dire sans altérer le spectre de la voix, et de manière non définitive.

⁸² Claudine Nougaret, Sophie Chiabaut, *Le son direct au cinéma, op. cit.*, p. 143.

Une alternative à ce problème est d'atténuer les fréquences basses en appliquant un filtre coupe-bas possédant une pente « douce », c'est-à-dire coupant les fréquences dans le bas du spectre à -3 voire -6 décibels (dB) par octave. En employant le filtrage de cette manière, il est alors possible de récupérer ces fréquences en post-production car ces dernières sont atténuées et non « effacées » dès la prise de son.

Durant ce même atelier, les discussions autour du filtrage à la prise ont permis d'obtenir le constat suivant : le filtrage ou l'égalisation est inhérent au système de diffusion du son, c'est-à-dire à l'utilisation du casque et notamment du *HD25* de chez Sennheiser. Ce casque permet de « *juger plus facilement de l'effet de présence d'une voix sur un microphone cravate ou une perche placée très près* »⁸³ car il est généreux dans les basses fréquences. « *Il est donc recommandé de limiter les corrections de filtrage passe-haut dès que l'on utilise ce casque* »⁸⁴. Néanmoins, il est convenu qu'au vu de la calibration des écoutes et du traitement acoustique des salles de montage image généralement mal réglée, il est préférable d'atténuer les basses fréquences du *mixdown*. Cette recommandation insinue que le *mixdown* n'est plus du tout exploitable (même le signal de la perche) lors du montage paroles. Les propos suivants permettent de synthétiser l'ensemble de ce débat : « *une solution : peu ou pas de filtrage sur les pistes iso enregistrées en mode post-fader et filtrage plus important sur le mixdown. En revanche cette solution est bancal dans le sens où il faudra prendre la décision de ne pas garder le mixdown pour le mixage final. Une discussion s'impose donc avec le monteur paroles et le mixeur avant de prendre en compte cette option.* »⁸⁵.

Après la préamplification du signal, le chef opérateur du son dispose d'autres traitements fréquentiels. Quelle soit analogique ou numérique, l'égalisation est présente pour corriger le spectre des signaux enregistrés. Or ces égalisations sont alors irrémédiables et aucunement référencées sous forme de métadonnées ou dans le rapport son. Il est tout à fait envisageable que le chef opérateur du son renseigne les paramètres d'égalisation utilisés dans la section *Note* afin de les informer sous forme de métadonnées. Le problème qui émerge de ce traitement est qu'il est réalisé avec un casque ne correspondant en aucun cas à la bande passante d'enceintes de cinéma.

⁸³ Lucien Balibar, *La chaîne du son au cinéma et à la télévision : De la prise de son à la post-production*, op. cit., p. 158.

⁸⁴ Thierry Lebon, « Compte-rendu - L'atelier « Le son direct » », op. cit.

⁸⁵ Thierry Lebon, *ibid.*

Dans la pratique :

En ce qui concerne la perche, le filtrage est délicat car elle fournit une continuité sur l'ensemble d'une séquence voire sur toute la longueur du film.

Au niveau des microphones cravates, il existe un filtrage « naturel » dû à son placement et à sa position dans les diverses couches de vêtements. De par le contrôle difficile des microphones cravates à cause de la peau du comédien et des vêtements, certains chefs opérateurs du son utilisent un filtrage « technologique » pour compenser cet effet. Par exemple, si un microphone cravate sonne trop *bouché*⁸⁶, comme dirait Pierre Gamet, alors à travers une égalisation, il peut redonner un peu plus de clarté à ce signal. Cette égalisation est réalisée dans le but de ne plus reconnaître le timbre spécifique du microphone cravate. Toutefois, « *il n'est pas question de prétendre s'attaquer en direct à un filtrage précis de lumière sifflante, de caméra soufflante ou de parasite électrique.* »⁸⁷.

II.3.2/ Traitements dynamiques

Dans les enregistreurs audionumériques, les traitements dynamiques correspondent aux limiteurs et à la modulation de volume. Les limiteurs sont employés en fonction d'un certain niveau du signal avant la CAN. Le seuil, niveau déterminé par l'utilisateur, est une valeur déclenchant le fonctionnement du limiteur. Ce traitement permet de ne pas saturer le signal au moment de l'enregistrement.

Le limiteur, en plus de se prémunir de la saturation, permet de « *réduire volontairement un peu la dynamique de la parole, conformément à ce qui sera effectué à l'aide de compresseurs en mixage* »⁸⁸. Ce traitement permet d'« *adoucir* » les crêtes du signal afin d'avoir « *une plage dynamique proche de celle exploitée en salle* »⁸⁹ et donc de faciliter le travail du mixeur.

La modulation de volume est réalisée avec les potentiomètres à glissière, appelées aussi *faders*. Ce traitement est notamment utilisé dans la fabrication du *mixdown*.

⁸⁶ Claudine Nougaret, Sophie Chiabaut, *Le son direct au cinéma, op. cit.*, p. 31.

⁸⁷ Thierry Lebon, « Compte-rendu - L'atelier « Le son direct » », *op. cit.*

⁸⁸ Vincent Magnier, « Compte-rendu « Exploiter le son direct du plateau à l'auditorium avec le nouveau Aaton Cantar X3 » », Association Française du Son à l'Image, février 2016, URL : <https://www.afsi.eu/articles/24659-le-nouveau-aaton-cantar-x3>

⁸⁹ Vincent Magnier, *ibid.*

À la question « Pourquoi moduler ? », les réponses sont multiples. L'une d'elle est avant tout d'ordre technique et correspond au fait de fournir un signal « calibré » situé entre deux zones pour une bonne exploitation du son en post-production. Historiquement, le signal doit alors être entre la zone de saturation rendant le son inexploitable et la zone, liée au format d'enregistrement, du souffle de l'appareil. Elle permet d'obtenir un niveau général cohérent sur l'ensemble du tournage. Elle est extrêmement importante car les *rushes* audio sont écoutés à de nombreuses reprises avant d'arriver en post-production son, que ce soit pour les projections de *rushes* ou lors du montage image.

Néanmoins, « moduler » ne se limite pas à un niveau électrique : c'est aussi la sensibilité d'écoute et les sensations sonores au moment de la prise du chef opérateur du son. « Moduler », c'est accompagner des nuances de jeu, offrir une écoute originale de la scène. La modulation offre donc une première orientation esthétique au son direct.

De plus, « moduler » ne s'applique pas qu'au mouvement des *faders* : il s'agit aussi de la distance entre le comédien et la perche. Comme le disait Antoine Bonfanti, « ton premier potentiomètre, c'est ta perche »⁹⁰. À ce sujet, Éric Devulder explique « qu'il aime bien que [le perchman] prévienne les problèmes de dynamique et module à la perche. Il ne faut pas simplement coller un micro au bord du cadre, sans tenir compte de ce que fait le comédien. Le perchman doit savoir anticiper sur les forte en éloignant le micro et se rapprocher instinctivement sur les chuchotements. »⁹¹.

« Si le perchman, en dirigeant le micro vers la bonne source, est en charge de la « mise au point » (on dit qu'un son timbré est « net »), le chef opérateur du son définit l' « exposition », le niveau général de la scène, en jouant le choix des micros, leur positionnement et leur intensité. »⁹² (Lucien Balibar)

Comment rendre la modulation imperceptible ? Lucien Balibar, chef opérateur du son, donne son point de vue :

« Pour moduler discrètement, sans que l'auditeur ne puisse sentir les variations de gain apportées par l'opérateur, il faut exploiter les caractéristiques de l'oreille. Ou plus exactement ses petites failles avec le principe de masquage : simultané et temporel. [...] Juste après une parole prononcée suffisamment fort, il est donc facile de changer le niveau avant que l'oreille ne se référence à nouveau sur le silence qui

⁹⁰ Vincent Magnier, *Guide de la prise de son pour l'image – Reportage, documentaire, fiction en radio et télévision 3^e édition*, Paris, Dunod, 2007, rééd. coll. Audio-Photo-Vidéo, 2015, p. 66.

⁹¹ Claudine Nougaret, Sophie Chiabaut, *Le son direct au cinéma*, op. cit., p. 32.

⁹² Lucien Balibar, *La chaîne du son au cinéma et à la télévision : De la prise de son à la post-production*, op. cit., p. 125 - 126.

suit. Et juste au moment où le comédien reprend la parole, on dispose aussi d'un court laps de temps pour ajuster le niveau. Ce temps pendant lequel l'oreille est en quelque sorte désorientée s'appelle le temps d'intégration. Il varie entre 50 et 100 ms selon l'intensité du signal »⁹³.

La modulation de volume remplit les fonctions suivantes :

- réduction de la dynamique dès la prise ;
- amélioration de l'intelligibilité des dialogues ;

« L'opérateur du son peut améliorer considérablement l'intelligibilité de certains comédiens en compensant cette articulation un peu faible par une modulation poussée particulièrement sur les derniers mots prononcés. »⁹⁴

- mélange des sources ;
- ouverture et fermeture des appoints et des microphones cravates.

Dans diverses situations, il est nécessaire de couper certaines sources comme pour le déplacement bruyant d'un acteur, le passage d'un véhicule, l'ouverture ou la fermeture de porte ou encore le bruit d'un accessoire de jeu. De plus, lorsqu'un comédien ne parle pas, son microphone peut être coupé ce qui permet d'éviter un certain nombre de bruits parasites qu'enregistre chacun des microphones cravates et donc d'améliorer le signal mélangé du *mixdown*. Par exemple, trois microphones cravates mélangés sur une piste génèrent plus de *bruit de fond* qu'un unique microphone cravate.

En réalisant un parallèle avec le mixage cinéma réalisé en auditorium, la modulation est une mise en lumière des mots. Les propos de Jean-Pierre Laforce renvoient à ces enjeux fondamentaux :

« En mixage, ce sont des interventions très minimalistes. Par exemple, vous avez une réplique, et puis dans la réplique il va y avoir une petite inflexion que vous allez accentuer. Et cette petite inflexion va ou peut mettre à jour complètement quelque chose qui était là mais pas tant que ça éclairé. C'est un peu comme des petits spots, tout d'un coup on éclaire mieux un endroit de la réplique. Et ce travail fait presque réplique par réplique dans le film va produire quelque chose qui va être mieux, plus

⁹³ Lucien Balibar, *La chaîne du son au cinéma et à la télévision : De la prise de son à la post-production*, op. cit., p. 134.

⁹⁴ Lucien Balibar, *ibid*, p. 136.

fortement et plus sensiblement formulé. Je joue avec les intonations de la voix avec le niveau et mes petits doigts. Le mixage c'est avant tout une affaire de niveau. »⁹⁵.

Toutefois, cette modulation est rendue de plus en plus difficile à cause de l'augmentation des sources :

« Lors de l'utilisation de plusieurs micros HF en plus de la perche indispensable, l'opérateur est contraint d'abandonner quasiment la modulation de celle-ci qui devient un micro d'air [...]. Chaque micro est ouvert à un niveau moyen, et est au mieux ouvert ou fermé. »⁹⁶

On peut répertorier trois méthodes différentes dans la modulation de volume chez les chefs opérateurs du son en France :

- *La modulation avec le gain de préamplification*

Cette première méthode, qualifiée d' « ancienne école », est directement héritée des techniques de prise de son des premiers enregistreurs magnétiques monophonique et stéréophonique. Ici le gain de préamplification définit la modulation. Les chefs opérateurs du son peuvent alors décider de déclencher les limiteurs de ces préamplificateurs afin d'adoucir, dès la prise, les crêtes du signal. Cette méthode de modulation se scinde en deux pratiques différentes : la première correspond à la modulation analogique où le chef opérateur du son utilise le potentiomètre rotatif de préamplification de la console de mixage analogique ; la seconde pratique correspond au monde numérique où le chef opérateur du son assigne, à travers une matrice numérique, un potentiomètre rotatif ou rectiligne au préamplificateur microphonique de l'étage d'entrée.

- *La modulation en post fader*

Cette seconde méthode est pratiquement équivalente à la première mais s'en distingue par le fait d'affecter un gain de préamplification microphonique fixe avant le début de la prise. Le chef opérateur du son utilise alors les *faders* pour faire varier le gain de la piste. On distingue plusieurs manières de réaliser cette modulation en *post fader*, soit avec les *faders* d'une console de mixage analogique, soit avec les *faders* ou les potentiomètres rotatifs des enregistreurs numériques.

⁹⁵ **Laure ADLER**, « Emission HORS-CHAMPS – Jean-Pierre LAFORCE », France Culture, mai 2013, URL : <https://www.franceculture.fr/emissions/hors-champs/jean-pierre-laforce>

⁹⁶ **Lucien Balibar**, *La chaîne du son au cinéma et à la télévision : De la prise de son à la post-production*, op. cit., p. 137.

Ces deux premières méthodes définissent un niveau définitif au signal dans la piste de *mixdown* et dans la piste divergée. Elles sont risquées et peuvent engendrer le non-enregistrement de certaines répliques. Ces deux méthodes peuvent induire des sons ratés dus à l'improvisation des comédiens ou à des ouvertures de *fader* tardives. En réponse à cette pratique, lors de l'atelier « Le son direct », l'AFSI a proposé de ne pas fermer complètement les pistes et de laisser une légère modulation résiduelle afin qu'elle soit utilisable en post-production en utilisant le gain disponible (+ 36 dB) dans les STAN.

- *La modulation en pré fader*

La modulation en *pré fader* est une méthode employée avec l'avènement des enregistreurs multipistes. Que ce soit en analogique ou en numérique, elle correspond au prélèvement du signal avant le gain appliqué par le *fader*. Dès lors, le niveau de la source dans la piste divergée est uniquement donné par la préamplification appliquée lors de l'étage d'entrée. Cette méthode permet d'avoir la modulation du chef opérateur du son uniquement dans le *mixdown*.

En travaillant en *pré fader*, même lorsque les comédiens ne parlent pas, les pistes assignées aux microphones cravates restent ouvertes durant toute la durée de l'enregistrement. Cette solution est confortable pour le chef opérateur du son et elle suppose une plus grande souplesse en phase de post-production. Par exemple, si le chef opérateur du son venait à être en retard sur un début de réplique, les premiers mots ne seraient pas disponibles dans le *mixdown* mais seraient présents dans les pistes divergées. Cette méthode sous-entend une forme de non maîtrise de la dynamique où les signaux des pistes divergées peuvent être enregistrés plusieurs décibels en-dessous du signal dans le *mixdown*. Selon Florent Lavallée, mixeur cinéma, « travailler à partir d'une perche non-modulée est dommageable. Ce n'est pas dans l'intérêt du mixeur »⁹⁷.

⁹⁷ Vincent MAGNIER, « Compte-rendu « Exploiter le son direct du plateau à l'auditorium avec le nouveau Aaton Cantar X3 Association Française du Son à l'Image, février 2016, URL : <https://www.afsi.eu/articles/24659-le-nouveau-aaton-cantar-x3>

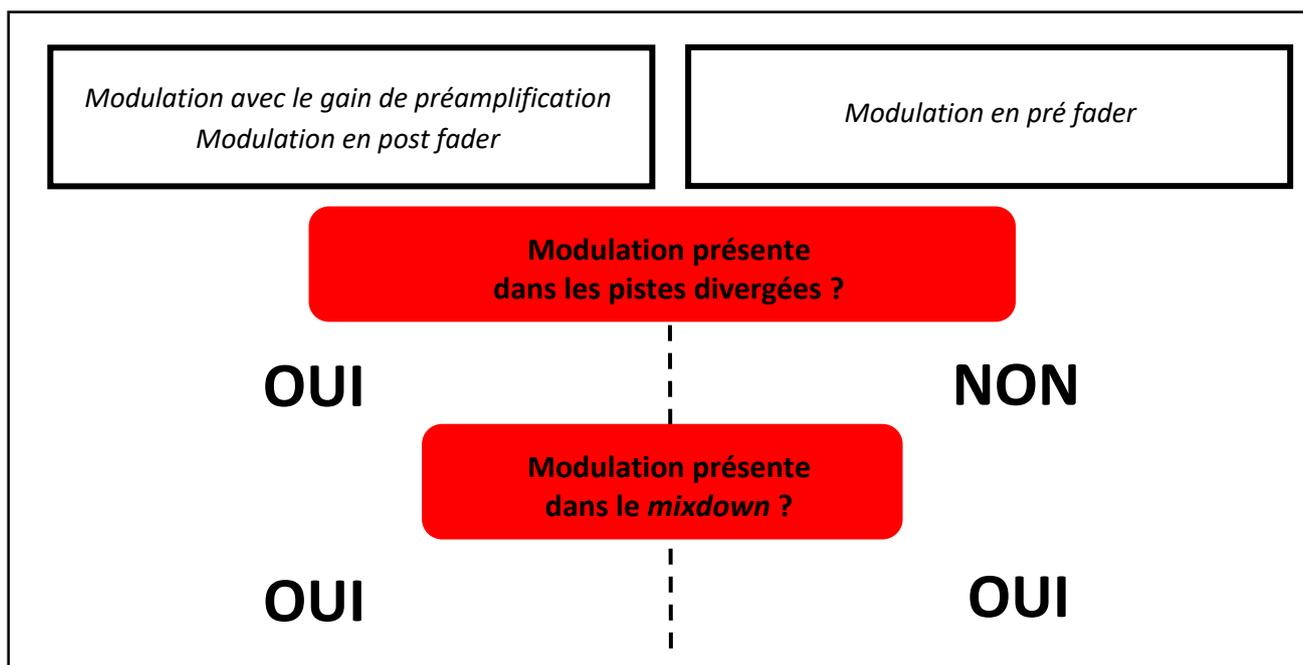


Figure 2.6 - Représentation des trois méthodes de modulation de volume

Finalement, on peut référencer deux manières de concevoir le travail au moment de l'enregistrement d'une prise de son au cinéma. La première est liée à l'héritage des premiers enregistreurs et notamment du mixage sur enregistreur bipiste. Comme pour le débat entre l'analogique et le numérique, il ne s'agit pas d'une confrontation entre « les passéistes d'un côté et les modernistes de l'autre : il s'agit plutôt de deux écoutes différentes »⁹⁸. Le chef opérateur du son propose une vision à travers les deux traitements présentés. Cette méthode ne se veut pas « diktat » comme l'on peut le supposer. Elle découle d'une pratique et d'un savoir-faire liés aux premières méthodes d'enregistrement. La seconde méthode est quant à elle, fortement liée aux technologies numériques. Sachant qu'il est désormais possible, grâce à des outils extrêmement pointus, de retravailler le son en post-production, l'ensemble de ces traitements ne sont faits qu'en cas de situations très spécifiques. Par exemple, les traitements fréquentiels se limitent uniquement au filtrage coupe-bas et la modulation est en *pré fader*. La prise de son se veut alors la plus neutre possible.

⁹⁸ Claudine Nougaret, Sophie Chiabaut, *Le son direct au cinéma, op. cit.*, p. 50.

On pourrait alors penser à l'emploi de sources livrées en deux versions : « modulée » sur une piste et fixe dite *pré fader* sur une autre piste. Néanmoins, cette méthode n'est pas préférable en post-production car elle multiplie par deux le nombre de pistes à traiter lors de la phase de montage paroles.

Quels sont les avantages et les inconvénients de la modulation de volume ?

- Avantages

La modulation est la sensibilité du chef opérateur du son

Elle correspond à l'écoute d'un individu proposant une retranscription tout à fait personnelle.

« J'aime bien parler d'empreintes, chaque ingénieur du son a forcément un coup de patte, une façon de faire. Heureusement, sinon le métier serait vraiment ennuyeux. Je sens les choses essentiellement avec mon oreille et à travers ce coup de patte que j'ai avec mon potentiomètre. »⁹⁹ (Pierre Gamet)

La modulation n'est pas un outil dangereux

La phase de tournage d'un film est en évolution permanente. Toutefois, le chef opérateur du son a la possibilité à travers les répétitions de s'exercer et de vérifier les moments précis où les comédiens « jouent » leur texte. Même si cette expérience est réalisée en direct, elle découle d'un savoir-faire et n'est pas faite au hasard.

La prise de son au cinéma ne correspond pas non plus à un « réseau de caméra de surveillance » : elle découle d'une réflexion et de choix définitifs dès la prise. Ne pas utiliser la modulation du chef opérateur du son, c'est devoir refaire un travail déjà réalisé dès la phase de tournage. Reporter des intentions, c'est augmenter considérablement le travail en aval. Qui souhaiterait travailler sur des *rushes* sans véritable choix artistique ? Au cours des entretiens avec les professionnels du son, il est apparu que travailler sur une bande son en *pré fader*, c'était comme devoir « ranger sa chambre » : un travail ingrat.

⁹⁹ Claudine Nougaret, Sophie Chiabaut, Le son direct au cinéma, *op. cit.*, p. 13.

- Inconvénients

Dispositifs de tournage de plus en plus complexes

Les séquences tournées à plusieurs caméras empêchent de connaître l'image utilisée au moment du montage. Il est alors difficile de réaliser un mixage avec des intentions fortes sans connaître la valeur finale du plan.

L'improvisation

Elle empêche indéniablement de moduler le volume en direct car le chef opérateur du son travaille sans repère.

Fermer une source définitivement

Dans certains cas précis, le choix entre deux microphones cravates est possible. Par exemple, lorsque deux comédiens sont face à face, la réplique du comédien A peut être meilleure dans le microphone cravate du comédien B.

Lors des échanges avec des monteurs parole, ces derniers regrettent parfois de ne pas pouvoir chercher de la matière dans les microphones cravates fermés.

Phase de post-production plus adéquate au visionnage et à l'écoute du film

En salle de montage, la taille de l'image est beaucoup plus grande que celle disponible en tournage. Ce simple facteur change considérablement notre vision de la séquence. Le système d'écoute est quant à lui plus proche de la diffusion finale du film. Enfin, les outils mis en œuvre ont la particularité d'être réversibles : le monteur a la possibilité de revenir en arrière et donc de faire autant de fois qu'il le souhaite un traitement.

Phase de montage image

Le montage image engendre de nombreuses modifications dans le film écrit sous forme de scénario : des répliques disparaissent, les débuts ou fins de séquence sautent ou encore un plan séquence peut être coupé en son milieu. Tous ces éléments déconstruisent les données enregistrées en tournage ce qui implique une souplesse indispensable des données « sources ».

En conclusion, capter la voix, c'est enregistrer des moments uniques. Le chef opérateur du son dispose d'équipements importants pour couvrir différentes situations sur le terrain. La complexité croissante des équipements d'enregistrement, à travers l'avènement du multipiste, soulève différents débats. Aujourd'hui, les outils d'enregistrement tendent à évoluer et l'ombre du *mixage automatique* est de plus en plus réelle. Par exemple, l'enregistreur NOMAD de Zaxcom embarque un module d'auto-mixage se présentant comme une aide lors de configurations lourdes. Sound Devices propose l'outil *Dugan automixing* implémenté dans l'enregistreur 688.

Peut-on enregistrer différemment les traitements réalisés en direct par le chef opérateur du son ?

À travers l'extrait de texte de Lucien Balibar présenté en *Annexe H*, la solution semble se trouver dans la virtualisation des procédés de traitement de tournage. C'est par ce cheminement que j'ai eu le désir de travailler sur de nouvelles métadonnées permettant de retranscrire le travail du chef opérateur du son.

Chapitre III

Enregistrement et exploitation de nouvelles métadonnées

À tout moment, l'utilisateur peut décider de supprimer ou modifier l'ensemble des métadonnées d'un fichier audio. Ainsi ce travail ne s'interroge pas sur l'archivage des fichiers audio. La proposition de ce mémoire concerne uniquement le « transport » d'informations provenant du tournage vers la phase de post-production. Elles sont spécifiques à une étape de travail au cinéma dans le cadre d'un flux de travail audionumérique.

Au regard des pratiques des chefs opérateurs du son, de nombreuses données du tournage ne sont pas exploitées sous forme de métadonnées. Par exemple, les traitements de modulation de volume, de filtrage et de phase sont appliqués directement et de manière définitive au signal mais ne sont pas référencés.

Dans ce troisième chapitre, nous présenterons dans un premier temps, de nouvelles métadonnées répondant à diverses problématiques soulevées au cours du *Chapitre II*. Ces données liées aux traitements du chef opérateur du son sont déjà employées quotidiennement par les professionnels. Une analyse des propositions et implémentations actuelles sur le marché permettra de définir chaque paramètre que l'on souhaite transposer en métadonnées. Dans un deuxième temps, nous proposerons une mise en forme de ces données et nous décrirons les différents enjeux que soulève la proposition de ce mémoire. Enfin, dans un troisième et dernier temps, nous présenterons le démonstrateur qui constitue la Partie Pratique du Mémoire (PPM) ainsi que ses objectifs et son développement.

III.1/ Présentation des nouvelles métadonnées

D'une part, au vu des pratiques des chefs opérateurs du son et des débats autour des méthodes d'enregistrement, ce mémoire propose une solution pratique en exploitant de nouvelles métadonnées. D'autre part, malgré le travail des fabricants et des professionnels, les nouvelles métadonnées ne référencent que des données majoritairement statiques. Aucune information sur le travail de fond du chef opérateur du son, c'est-à-dire sur les traitements fréquentiels et dynamiques ne sont transmises à la post-production. Le chef opérateur du son ne fournit qu'une partie des informations.

Par la lecture des mémoires de fin d'études de Baptiste Mésange, « Dimensionnement et implémentation d'un enregistreur portable modulaire »¹⁰⁰ (2017) et de Dorian Racine, « Les microphones numériques et la prise de son à l'image »¹⁰¹ (2016) mais aussi par la lecture du livre de Lucien Balibar, *La chaîne du son au cinéma et à la télévision*¹⁰², m'est venue l'idée de proposer de nouvelles métadonnées répondant différentes problématiques de tournage.

III.1.1/ Définitions

Les métadonnées dynamiques renvoient à des paramètres fluctuant dans le temps. Elles sont ainsi indissociables d'une donnée temporelle. Lorsqu'un utilisateur réalise différents traitements dans une STAN, il associe la valeur d'un paramètre à une donnée temporelle. Par exemple, lorsqu'il réalise une automatisation de volume, la Station de Travail Audionumérique (STAN) enregistre la position du potentiomètre sous forme de points à une position temporelle fixe. Au vu des traitements disponibles dans les consoles de mixage analogiques et les enregistreurs audionumériques actuels, ces traitements sont d'ordre dynamique, fréquentiel, de phase et de panoramique.

Il faut noter que seuls les paramètres situés après la Conversion Analogique-Numérique (CAN) sont uniquement destinés à être mis sous forme de métadonnées. Toute intervention analogique en amont de la CAN ne peut être référencée par la suite en métadonnées et est appliquée de manière permanente au signal. Ces interventions sont dites **destructives**¹⁰³. Voici deux situations différentes :

Le chef opérateur du son utilise une console de mixage analogique

La console de mixage analogique se situe en amont de l'enregistreur audionumérique. L'ensemble de ces traitements est appliqué au signal en amont de la CAN réalisée dans l'enregistreur.

¹⁰⁰ **Baptiste Mésange**, Baptiste, MUDO : Dimensionnement et implémentation d'un enregistreur portable modulaire, mémoire sous la direction de Sylvain LAMBINET et Mattéo CASSONNET, École Nationale Supérieure Louis Lumière, section son, 2017.

¹⁰¹ **Dorian Racine**, Les microphones numériques et la prise de son à l'image, mémoire sous la direction de Frédéric BELIN et Axel BRISARD, Université d'Aix-Marseille, Département Sciences, Arts et Techniques de l'Image et du son, 2016, p. 53.

¹⁰² **Lucien Balibar**, *La chaîne du son au cinéma et à la télévision* : De la prise de son à la post-production, *op. cit.*

¹⁰³ Le terme « destructif » réfère à toutes les actions qui empêchent un retour en arrière. Inversement, le terme « non destructif » renvoie à toutes les opérations dont l'utilisateur possède un historique. Par exemple, dans les STAN, l'utilisateur a accès aux lignes d'automatisation de ses traitements en insert de piste et peut modifier à tout moment un paramètre.

Le chef opérateur du son utilise uniquement un enregistreur audionumérique

Même dans cette situation, le chef opérateur du son réalise différents traitements analogiques avant la CAN. Par exemple, le filtrage coupe-bas est appliqué au signal de manière définitive. Ensuite, les autres traitements sont réalisés après dans le domaine numérique.

Les quatre nouvelles métadonnées proposées sont :

- filtrage ;
 - panoramique ;
 - modulation de volume ;
 - phase.
-
- Métadonnées statiques

Lors de l'enregistrement du phénomène sonore, le chef opérateur du son réalise des traitements fréquentiels et une affection, ou plutôt un panoramique, des sources afin de créer le *mixdown*. Réalisés en début de prise, en général ces traitements n'évoluent pas au cours du temps. Toutefois, de par l'implémentation actuelle des enregistreurs, il est possible de les modifier en cours de prise. Certaines de ces données sont déjà existantes sous forme de métadonnées dans des champs privés du iXML. Par exemple, Aaton renseigne déjà le filtrage coupe-bas et le panoramique dans les métadonnées générées par le Cantar X3. Au regard des pratiques actuelles, les données de traitements fréquentiels et de panoramique à mettre sous forme de métadonnées sont :

Filtrage et égalisation

Afin de caractériser un filtre, les données utiles sont : la pente, exprimée en décibels par octave, et la fréquence de coupure (F_c) en hertz.

L'égalisation regroupe plusieurs types de filtre : le filtre baxendall, le filtre semi-paramétrique, le filtre paramétrique et les filtres à plateau (Shelf). Pour caractériser les trois premiers, il faut disposer des données suivantes :

- *efficacité (en décibels) correspondant à une amplification ou à une atténuation ;*
- *fréquence centrale (F_c) variable ;*
- *sélectivité (Q).*

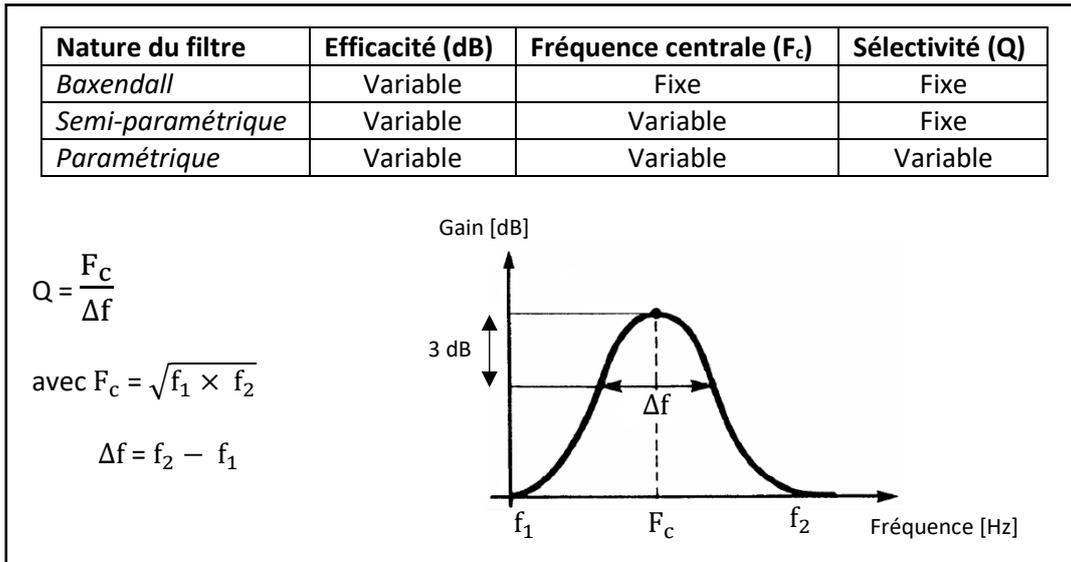


Figure 3.1 - Présentation du filtrage Baxendall, semi-paramétrique et paramétrique

Enfin, les filtres Shelf se caractérisent par une fréquence de coupure (F_c) et un gain.

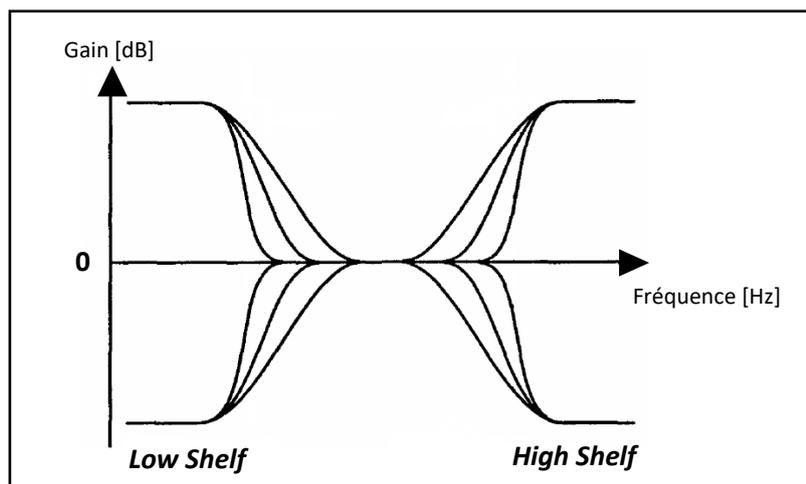


Figure 3.2 - Présentation des filtres à plateau
(à gauche Low Shelf ; à droite High Shelf)

Panoramique

L'emploi du panoramique en tournage permet d'affecter le signal d'une piste divergée à une des deux pistes du mixdown. Cette valeur est exprimée en pourcentage.

- Métadonnées dynamiques

Elles correspondent à des traitements évoluant dans le temps effectués sur le signal. Par exemple, la modulation de volume et l'opposition de phase sont les traitements dynamiques dont dispose le chef opérateur du son.

Modulation de volume

Le mouvement de fader est l'élément clé de la modulation de volume. Réalisé avec les potentiomètres rotatifs ou à glissière (fader), sa position évolue constamment lors d'un enregistrement en fonction de la source sonore.

Opposition de phase

L'opposition de phase correspond à un traitement permettant de réaliser un déphasage de 180° du signal. D'un point de vue électronique, l'opposition de phase est représentée par le signe « Ø » et correspond à une inversion de polarité.

III.1.2/ Où se situent ces nouvelles métadonnées dans la chaîne du son?

Aujourd'hui, dans les enregistreurs audionumériques, les quatre traitements présentés ci-dessus sont pratiquement tous réalisés après la CAN, c'est-à-dire dans le domaine numérique. Ils sont donc éligibles à devenir des métadonnées. À noter que le filtrage coupe-bas est analogique mais peut être tout de même informé dans les métadonnées. L'extraction de ces métadonnées se place au niveau des flèches jaunes de la *Figure 3.3*. Ce schéma prend en considération l'*Annexe G* qui présente la chaîne du son dans un Cantar-X3 d'Aaton.

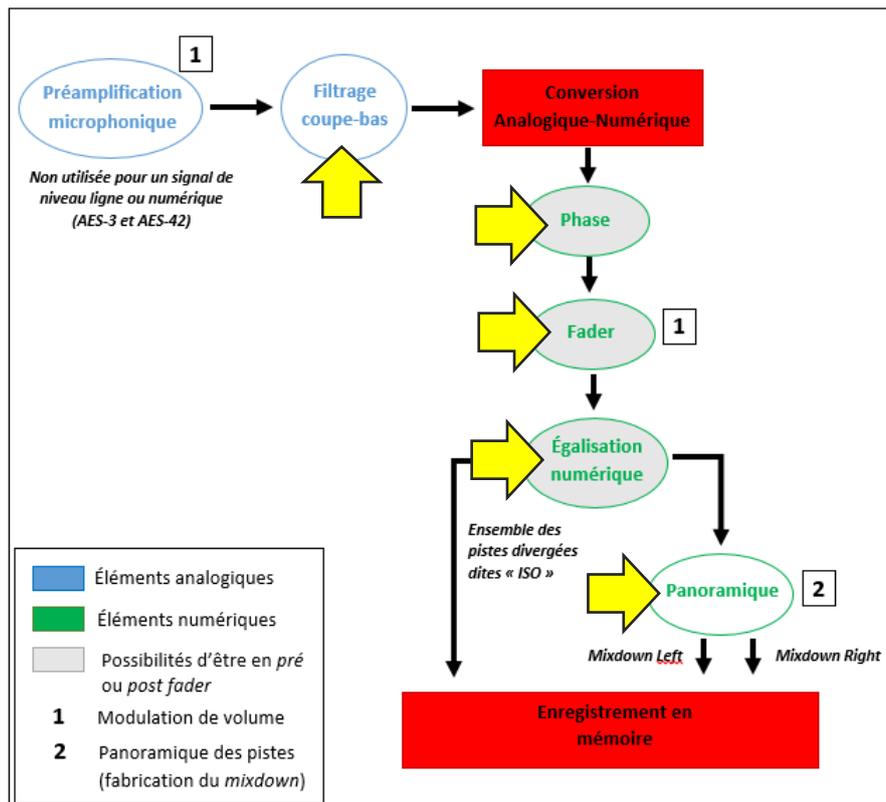


Figure 3.3 - Schéma représentant l'emplacement des données dans un enregistreur audionumérique

Tous les paramètres des traitements à mettre sous forme de métadonnées se situent entre la préamplification du signal et l'enregistrement de l'ensemble des pistes sur un support disque dur (type SSD) au format BWF. Il faut noter que la position de ces traitements varie en fonction des intégrations matérielles et logicielles des constructeurs d'enregistreur.

III.1.3/ Analyse des paramètres des nouvelles métadonnées

Dans cette partie, nous analyserons successivement les paramètres de chaque traitement à récupérer dans l'ordre suivant :

- modulation de volume ;
- opposition de phase ;
- filtrage ;
- panoramique.

Les différentes valeurs obtenues pour chaque paramètre des traitements ci-dessus découlent d'une recherche dans la documentation des constructeurs, dans les spécifications liées aux formats audio existants et dans les informations recueillis dans divers forums liés au son.

Premières métadonnées : modulation de volume

La modulation correspond à une métadonnée dynamique. De par sa nature, elle comprend deux paramètres intimement liés. Le premier paramètre est la position du curseur qui donne une atténuation ou non du volume en décibels. Comme pour la numérisation du son sous la forme PCM, on réalise un échantillonnage et une quantification du signal analogique afin de lui attribuer un « format » numérique. Le second paramètre est quant à lui lié à une donnée temporelle. Dans le cadre de la numérisation de la position de *fader*, il faut procéder à une numérisation du signal. Les paramètres à définir pour la modulation de volume sont donc :

- la fréquence d'échantillonnage : F_sF ;
- la valeur délivrée par le potentiomètre quantifiée.

Fréquence d'échantillonnage du fader (F_sF)

Le résultat de la valeur de F_sF résulte d'une phase de tests successifs réalisés dans Pro Tools (version 12.7.1). Dans cette STAN, différents paramètres liés à l'automatisation de volume sont configurables par l'utilisateur dans le menu *Préférences-Mixage*. L'utilisateur peut alors activer ou non le paramètre « *Affiner et réduire les données après la passe*¹⁰⁴ » et lui affecter la valeur suivante :

- aucun(e) ;
- petit ;
- moyen ;
- plus ;
- le plus.

Ce paramètre correspond à la suppression ou non de points d'automatisation de manière automatique afin de réaliser un lissage des données et de supprimer les points redondants, donc inutiles, enregistrés temporairement lors de la passe. Nous avons réalisé l'enregistrement de cinq passes correspondant aux cinq possibilités de réduction de points en effectuant des mouvements de *faders* chaotiques et

¹⁰⁴ Le terme « passe » est un mot employé dans le vocabulaire du mixage pour désigner l'ensemble des actions suivantes : en premier lieu l'action de lancer la lecture dans la STAN puis de réaliser les traitements de manière statiques et/ou dynamiques et enfin d'arrêter la lecture en appuyant sur Pause.

non représentatifs d'une modulation de volume « naturelle ». Les courbes obtenues exposées en *Figure 3.4* ci-dessous ont permis de « calculer » une valeur expérimentale de FsF.

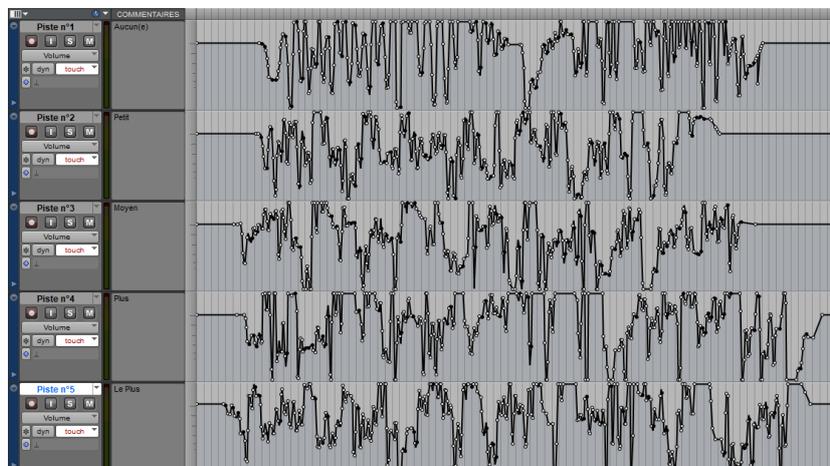


Figure 3.4 - Capture d'écran de la recherche de FsF dans Pro Tools

En utilisant l'outil d'édition « Sélecteur », nous avons calculé la distance en nombre d'échantillons séparant deux points d'automation.

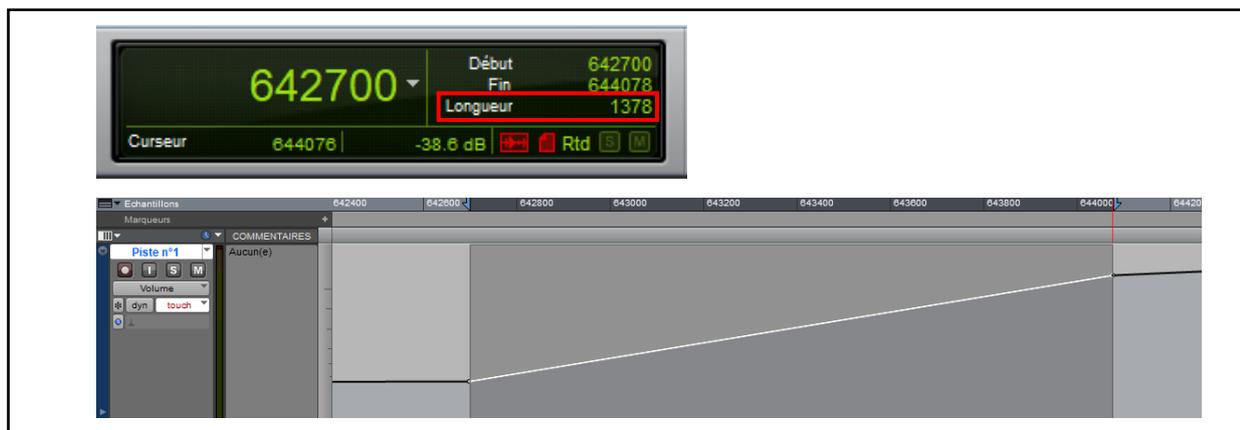


Figure 3.5 - Capture d'écran représentant le calcul séparant deux points d'automation dans Pro Tools

Après répétition de cette manipulation, la distance « moyenne » séparant deux points d'automation quelconque est approximativement autour de 1400 échantillons (dans la *Figure 3.5* ci-dessus, la distance entre deux points d'automation est de 1378 échantillons).

Que représentent ces 1400 échantillons ? À une fréquence d'échantillonnage de 48 kHz, cette valeur correspond à 39 ms. Selon Paul Charlemagne, le « temps d'intégration, ou encore seuil de durée *interstimuli*, se situe entre 25 et 50 ms. C'est l'intervalle de temps nécessaire pour que les stimuli

sonores soient perçus comme successifs et non simultanés : en dessous de ce seuil, les stimuli fusionnent. »¹⁰⁵. Ces 39 ms correspondent donc justement à une valeur située au niveau du temps d'intégration de l'oreille.

On retrouve aussi la notion de temps d'intégration dans les propos de Lucien Balibar : « *Au moment où le comédien reprend la parole, on dispose aussi d'un court laps de temps pour ajuster le niveau. Ce temps pendant lequel l'oreille est en quelque sorte désorientée s'appelle le temps d'intégration. Il varie entre 50 et 100 ms selon l'intensité du signal.* »¹⁰⁶. Au vu de la pertinence entre la valeur temporelle expérimentale et la valeur théorique liée au fonctionnement physique de l'oreille, nous décidons de travailler à partir de cette valeur. Dès lors, cette distance de 1400 échantillons sera utilisée comme la distance séparant tous les points d'automation entre eux.

Cette donnée temporelle suppose d'enregistrer la position du *fader* tous les 1400 échantillons. Or, dans la pratique, le chef opérateur du son ne module pas tout le temps. Nous avons donc réfléchi à une seconde méthode indépendante d'une fréquence d'échantillonnage.

Nous pouvons généralement lier une valeur d'un paramètre à une valeur temporelle *absolue* ou *relative*. La valeur temporelle absolue correspond au *Timecode*. Celui-ci permet de subdiviser la seconde en 24 points. Après un rapide calcul, nous obtenons que le *Timecode* permet d'avoir une précision entre deux positions temporelles différentes de 42 ms. Il est alors tout à fait envisageable de l'employer comme solution. Malgré son aspect absolu, il n'est pas utile d'avoir l'heure à laquelle le point d'automation a été créé. De plus, la lourdeur de la norme *Timecode* nous pousse à utiliser une valeur temporelle *relative* : la *position temporelle en échantillons*. Cette valeur relative correspond au nombre d'échantillons écoulés depuis le début de l'enregistrement. Cette dernière proposition est conservée dans le cahier des charges du démonstrateur.

¹⁰⁵ Paul Charlemagne, *Psychoacoustique : Mesure subjective des paramètres d'intensité, de fréquence et de temps*, mémoire sous la direction de Yves LASRY, Université Henri Poincaré – Nancy Université, Faculté de Pharmacie, 2010, p. 25.

¹⁰⁶ Lucien Balibar, *La chaîne du son au cinéma et à la télévision : De la prise de son à la post-production*, *op. cit.*, p. 134.

Calcul du nombre de bits alloués pour une valeur temporelle relative

Dans le cadre de l'utilisation d'une fréquence d'échantillonnage de 48 kHz, le calcul est le suivant :

48 000 (échantillons) x 60 (seconde) x 60 (minute) x 24 (heures) = 4 147 200 000 valeurs

$2^{32} = 4\,294\,967\,296$ valeurs

$2^{31} = 2\,147\,483\,648$ valeurs

Il faut donc disposer de 32 bits pour enregistrer la valeur temporelle *relative* exprimée en nombre d'échantillons.

Figure 3.6 - Calcul du nombre de bits alloués pour la valeur temporelle relative

Quantification appliquée au signal délivré par le fader

Dans la pratique, les potentiomètres à glissière et les potentiomètres numériques délivrent une tension en sortie qui est convertie en une valeur numérique et appliquée au signal en temps réel. Au final, le signal numérisé est atténué ou augmenté d'un certain gain en décibels. Afin de déterminer une valeur pertinente de quantification pour les *faders*, les recherches bibliographiques se sont portées sur deux sources disponibles : la littérature au sujet du nombre de bits alloués à la position du *fader* et les caractéristiques physiques des potentiomètres actuels employés dans les surfaces de contrôle et les consoles de mixage.

Il existe plusieurs manières d'écrire l'automation de modulation de volume. Les recherches se sont donc tournées vers les différents protocoles de communication spécifique à l'audio et les fichiers audionumériques incorporant des données liées à la modulation de volume. Voici une liste des ressources consultées :

- station de Travail Audionumérique (STAN) ;
- protocoles HUI et Logic Control ;
- consoles numériques et surface de contrôle ;
- format AAF ;
- fichier OMF ;
- fichier MIDI.

Concernant les STAN, il s'est avéré impossible de trouver des ressources bibliographiques spécifiques à ce sujet car ces informations sont propriétaires.

Le protocole HUI a été développé en 1997 par les sociétés Digidesign et Mackie. Il n'est pas documenté car il est propriétaire. Toutefois, en 2010, un membre de la communauté de Reaper a réalisé la rétro-ingénierie de ce protocole et a publié les résultats¹⁰⁷. La résolution totale de la valeur d'un *fader* est de 14 bits. Or, selon ses recherches, la résolution intrinsèque à un *fader* quelconque est de 9 bits ce qui signifie que les cinq derniers bits les moins significatifs sont mis à 0 automatiquement.

D'après l'ensemble des propos lus sur les différents forums¹⁰⁸, une grande majorité des consoles et de surfaces de contrôle intègrent une quantification sur 10 bits.

L'*Advanced Authoring Format* (AAF) est bien documenté. Le document intitulé « *Advanced Authoring Format - Edit Protocol* »¹⁰⁹ témoigne d'une quantification de 31 bits pour le gain. On retrouve cette même valeur pour l'*Open Media Framework* (OMF) dans les forums.

En ce qui concerne le protocole MIDI, il est extrêmement bien documenté. La valeur du *Channel Volume* est contenue dans les *Control Change* n°7 (CC#7) et n°39. . En combinant ces deux *Control Change*, l'utilisateur accède donc à une résolution de 14 bits soit 16 384 valeurs au lieu des 128. Toutefois, la plupart des machines n'utilisent que le n°7. Dans la spécification du MIDI¹¹⁰, l'affectation d'une valeur MIDI à un gain en décibels est donnée par l'équation suivante :

$$L \text{ (dB)} = 40 \log_{10} (V/127) \text{ où } V = \text{CC\#7}$$

Ce qui donne en pratique :

127 = 0 dB	32 = - 23,9 dB
96 = - 4,9 dB	16 = - 36,0 dB
64 = - 11,9 dB	0 = - ∞

Au vu de l'ensemble des informations disponibles, nous décidons de fixer la quantification d'un *fader* à 12 bits.

Les potentiomètres à glissière actuels des surfaces de contrôle et des consoles de mixage analogiques mesurent entre 100 et 120 millimètres (mm). En employant un *fader* dont la position est quantifiée

¹⁰⁷ Consultable à l'adresse URL :

<https://forum.cockos.com/showthread.php?t=101328>

¹⁰⁸ Consultables aux adresses URL suivantes :

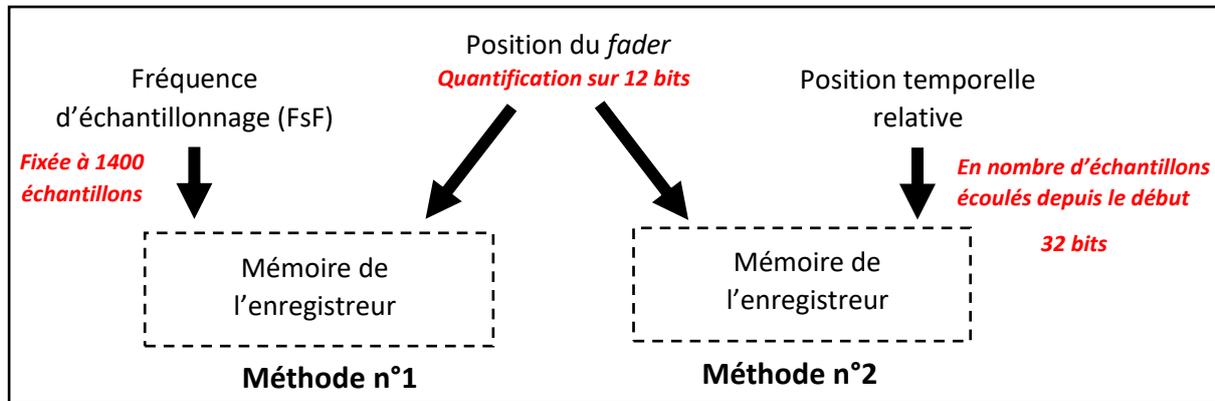
<http://duc.avid.com/> ; <https://www.gearslutz.com/> ; <https://forums.cockos.com/>

¹⁰⁹ **Advanced Media Workflow Association**, Spécification Advanced Authoring Format, Advanced Authoring Format (AAF) - Edit Protocol, avril 2008, p. 45 - 46.

¹¹⁰ **MIDI Manufacturers Association**, Spécification Musical Instrument Digital Interface, General MIDI 2, Version 1.2a, février 2007, p. 7.

sur 12 bits, il nous est possible de récupérer 34 points différents sur 1 mm. Cette quantification est donc largement suffisante au vu des données possiblement créées par l'utilisateur.

En termes d'implémentation, les deux propositions possibles sont présentées dans la *Figure 3.7* ci-dessous.



Métadonnées	Paramètre 1		Paramètre 2	
	Nom	Valeur	Nom	Valeur
Modulation de <i>fader</i> (méthode n°1)	Quantification	(12 bits)	FsF	1400 échantillons
Modulation de <i>fader</i> (méthode n°2)		de - ∞ à + 12 dB	Valeur Temporelle relative	En nombre d'échantillons depuis le début de l'enregistrement

Figure 3.7 – Présentation des implémentations possibles pour la modulation de volume

Deuxièmes métadonnées : le filtrage

Concernant le filtrage, il est envisagé qu'à chaque filtrage ou égaliseur paramétrique employé par l'utilisateur, différents paramètres sont conservés en mémoire afin d'être retranscrits sous forme de métadonnées. Sachant que le filtrage est une potentielle métadonnée statique, il n'est pas nécessaire de référencer une quelconque donnée temporelle à ces paramètres. Le choix des valeurs des paramètres liés au filtrage découle de l'analyse de l'ensemble des enregistreurs numériques et des consoles de mixage analogiques existantes. L'ensemble des informations sont synthétisées en *Annexe I*.

Dans un souci de compatibilité avec les solutions de mixage et d'édition proposées dans les STAN, nous avons analysé le *plug-in*¹¹¹ *Channel Strip*, employé dans le logiciel *Pro Tools*. La *Figure 3.8* ci-dessous reprend donc la synthèse réalisée à travers la lecture des documentations constructeurs et des paramètres disponibles dans le *Channel Strip*.

¹¹¹ Un *plug-in* est un programme extérieur qui interagit avec un programme principal nommé « hôte ». Ne pouvant fonctionner seul, il permet d'apporter de nouvelles fonctionnalités. Dans le domaine du son, il permet typiquement d'utiliser un instrument virtuel ou un effet audio.

Métadonnées	Paramètre 1		Paramètre 2		Paramètre 3		Paramètre 4	
	Nature		Nom	Valeur	Nom	Valeur	Nom	Valeur
Filtre	Passe-haut	Passe-bas	Fréquence de coupure (Fc)	0 à 20 000 Hz	Pente	- 3 à 24 db/octave	Gain	
	Low-Shelf	High-Shelf						+/- 24 dB
EQ	Nature		Fréquence centrale (Fc)	0 à 20 000 Hz	Q	0,1 à 100	Gain	+/- 24 dB
	Peaking							

Figure 3.8 - Tableau des paramètres de métadonnées de filtrage

Troisièmes métadonnées : l'opposition de phase

La phase ne nécessite pas une analyse des propositions des fabricants. C'est une valeur booléenne qui comprend deux états : OUI ou NON. Afin d'être employée sous forme de métadonnées dynamiques, il faut lui ajouter une donnée temporelle. Celle-ci peut être absolue ou relative. Comme pour la modulation de volume, nous décidons d'employer une valeur temporelle relative exprimées en nombre d'échantillons échantillons.

Métadonnées	Paramètre 1		Paramètre 2	
	Nom	Valeur	Nom	Valeur
Opposition de Phase	État	Non (0) ou Oui (1)	Valeur Temporelle relative	En nombre d'échantillons depuis le début de l'enregistrement

Figure 3.9 - Tableau des paramètres des métadonnées liées à la phase

Quatrièmes métadonnées : le panoramique

Le panoramique lié à la fabrication du *mixdown* est une métadonnée statique. Aucune valeur temporelle n'est donc nécessaire. Il suffit que la machine renseigne automatiquement la matrice d'affectation des pistes au *mixdown* en enregistrant l'état du panoramique en début ou fin d'enregistrement.

Métadonnées	Paramètre 1	
	Piste concernée	Valeur
Panoramique	Mixdown	Numéro des pistes affectées
	Piste divergée	Nombre entre 0 et 100*

(*)
Un couple de deux nombres est donné (X ; Y). Par exemple, si X est égal à 100 et Y à 0 alors le signal de la piste divergée est envoyé uniquement au *mixdown* Gauche.

Figure 3.10 - Tableau des paramètres des métadonnées liées au panoramique

III.2/ Exploitation de ces nouvelles métadonnées

III.2.1/ Mise en forme

La mise en forme de ces nouvelles métadonnées est primordiale. Elle doit s'inscrire dans le contexte actuel et répondre à une simplicité de programmation et d'intégration. La première proposition serait de présenter ces métadonnées dans une nouvelle tranche et de les mettre en forme comme dans la tranche BEXT avec des données binaires. Ci-dessous, un exemple de mise en forme possible de la première proposition :

- descripteur : **74 72 74 6D** (= trtm en codage ASCII) ;
- taille : exprimée en octets ;
- données de la tranche trtm : métadonnées mises en forme en hexadécimal.

Les métadonnées seraient présentées dans une mise en forme contiguë comme dans la tranche BEXT.

Néanmoins, cette première proposition apparaît complexe et cette mise en forme des données semble difficile à mettre en place. C'est pourquoi nous décidons d'utiliser le langage XML pour mettre en forme l'ensemble des nouvelles métadonnées présentées en début de chapitre. Les avantages du langage XML sont :

- une « grammaire » stricte et standardisée ;
- une structure hiérarchique permettant d'ordonner par ordre d'importance les informations ;
- l'utilisation du fichier texte ;
- un code verbeux et lisible directement par un être humain.

Le choix de ce langage s'inscrit aussi dans son utilisation faite par la tranche iXML. En reprenant les différents paramètres définis en *Partie III.1.3*, une mise en forme des métadonnées dans le langage XML est proposée en *Figure 3.11*. Dans un souci de compréhensibilité, les balises XML sont nommées en langue française.

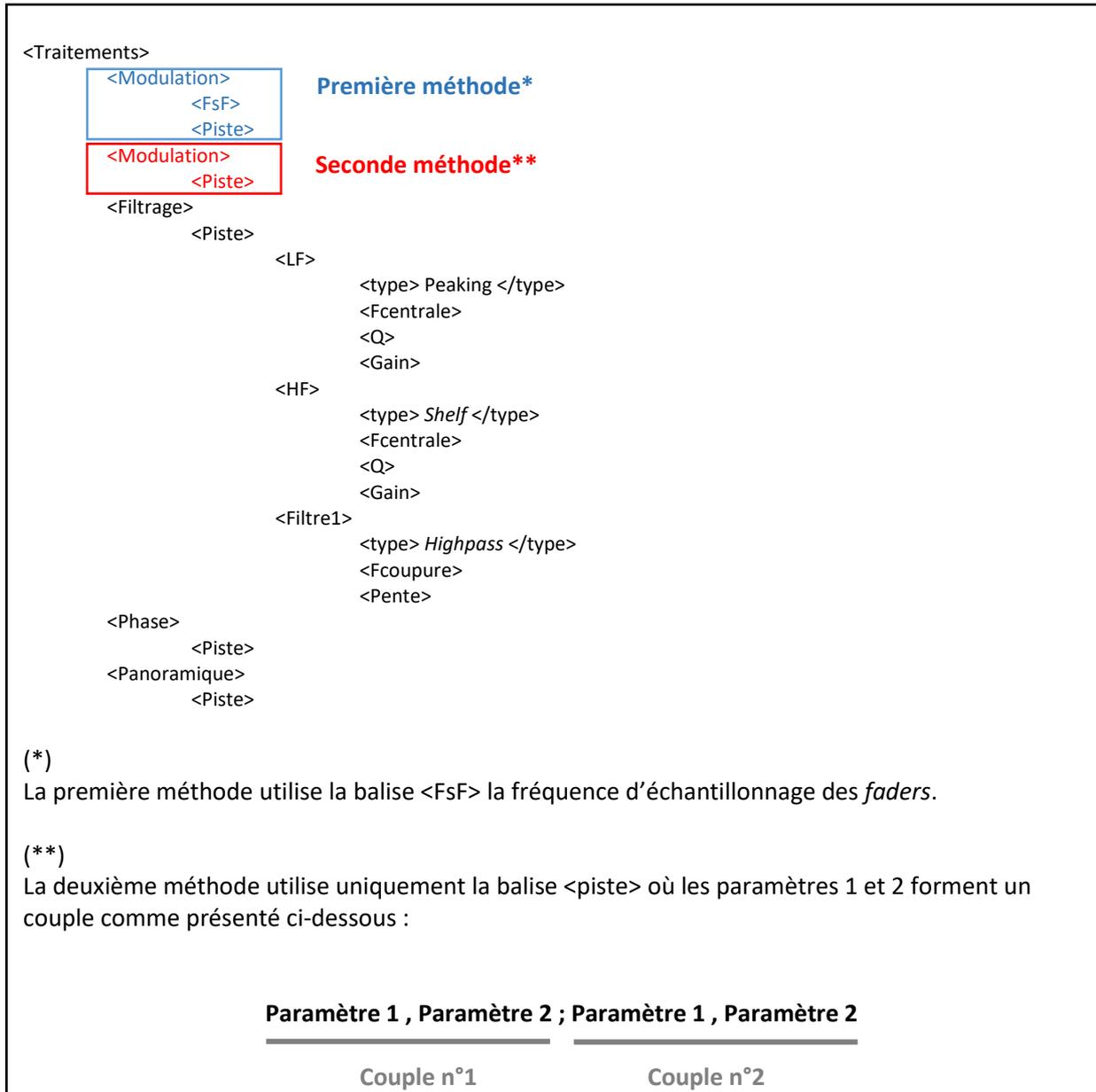


Figure 3.11 - Mise en forme en langage XML appliquée aux nouvelles métadonnées

Un exemple de fichier XML généré par le démonstrateur est présenté en *Annexe J*. Ce fichier texte reprend l'arborescence XML présentée ci-dessus : il correspond à la fabrication de toutes les métadonnées à partir de deux pistes où la première méthode (avec FsF) est appliquée pour la modulation de volume. L'*Annexe K* correspond à un autre fichier XML généré à partir du démonstrateur en appliquant cette fois-ci la seconde méthode pour les métadonnées de modulation de volume.

La disposition des balises XML utilisée correspond à un ordonnancement semblable aux tranches binaires comme le BEXT ou la tranche DONNÉES. Toutes les métadonnées dynamiques (modulation de

volume et de phase) se composent d'un couple répété indéfiniment en fonction de la durée de l'enregistrement. En d'autres termes, les informations se suivent et sont toutes disposées dans les mêmes balises. C'est une version optimale en termes de compacité. Mais cette mise en forme paraît peu évolutive. Si on désire implémenter un nouveau paramètre à l'une des métadonnées dynamiques, alors toute la structure est obsolète. Par exemple, au lieu de vouloir connaître la position temporelle relative en nombre d'échantillons, on désire implémenter la position temporelle absolue (*Timecode*), il faut revoir toute la structure du fichier et les outils permettant d'enregistrer et de lire les informations contenues dans ces mêmes balises.

L'objectif premier de la tranche iXML est de développer des métadonnées pérennes dans le temps. Ainsi, tant dans sa mise en forme que dans sa structure, un seul paramètre est affecté à une balise. En reprenant le schéma des balises <SYNC_POINT_LIST> et <TRACK_LIST> de la norme iXML, voici un développement possible des métadonnées dynamiques en *Figure 3.12* :

<VOLUME_LISTE>	<i>Cette balise indique le début des métadonnées concernant la modulation de volume</i>
<VOLUME_VALEUR_FsF>	<i>Cette balise indique la valeur de fréquence d'échantillonnage FsF</i>
< VOLUME_NOMBRE>	<i>Cette balise indique les positions de fader enregistrées (Nombre de points d'automation)</i>
<VOLUME>	<i>Cette balise regroupe les informations d'un unique point d'automation</i>
< VOLUME_VALEUR>	<i>Cette balise indique le gain appliqué au signal</i>
< VOLUME_LOW>	<i>Les deux dernières balises indiquent le nombre d'échantillons écoulés depuis le début de l'enregistrement</i>
< VOLUME_HIGH>	

Figure 3.12 - Évolution de la mise en forme des nouvelles métadonnées dynamiques

Reprenant les « canons » du iXML, cette mise en forme pose de réelles questions au niveau du poids du fichier. Il faudrait réaliser différents tests en comparant une mise en forme semblable à la tranche BEXT et une mise en forme reprenant les canons du iXML. Toutefois, cette structure semble bien adaptée pour la métadonnée dynamique de phase car l'utilisateur ne crée que peu de points par prise. La question porte surtout sur les métadonnées dynamiques de modulation de volume.

III.2.2/ Question du poids

La première question soulevée par les différents professionnels sur l'implémentation de ces nouvelles métadonnées concernait le poids ! C'est pourquoi, tout au long de la mise en œuvre du projet, nous avons tâché de créer un fichier texte externe au fichier audio, afin de « peser » directement ces nouvelles métadonnées.

Au moment de passer sur la « balance », la question porte surtout sur les métadonnées dynamiques (modulation de volume et phase). La phase n'est utilisée que très rarement, c'est pourquoi toute l'attention est portée sur les métadonnées de modulation de volume. Lors des premières expérimentations, les résultats obtenus par calculs théoriques sont présentés en *Figure 3.13* ci-dessous.

On possède les éléments suivants :

- Nombre d'échantillons = 4 147 200 200 = 32 bits = 40 bits (en UTF-8) ;
- Echantillonnage fixé (pour le moment) = 1400 échantillons ;
- Quantification = 12 bits (de $-\infty$ à + 12 dB) ;
Pour coder « -140,0 » en UTF-8, il faut 20 bits
- Durée du fichier : 3 minutes (180 secondes).

Calculs en prenant en compte un codage en UTF-8 (XML)

Méthode n°1

Taille = $180 \times 48\,000 / 1400 \times 20 = 15,5$ kilo-octets [ko]

Méthode n°2 (ici le calcul est faussé car nous supposons que le chef opérateur du son bouge constamment la position du *fader*)

Taille = $180 \times 48\,000 / 1400 \times (20+40) = 46$ ko

Calculs en prenant en compte un codage binaire simple

Méthode n°1

Taille = $180 \times 48\,000 / 1400 \times 12 = 9$ ko

Méthode n°2 (encore une fois le calcul est faussé pour les mêmes raisons citées ci-dessus)

Taille = $180 \times 48\,000 / 1400 \times (12+32) = 33$ ko

Figure 3.13 - Calculs théoriques du poids des nouvelles métadonnées dynamiques (modulation de volume)

Au regard des résultats théoriques obtenus, le poids ne semble pas une question cruciale. On pourrait considérer que les métadonnées ont l'avantage de prendre peu d'espace (quelques kilooctets). Le problème était de devoir choisir entre la taille du fichier ou la précision des paramètres des métadonnées enregistrées. Cependant, au cours des échanges avec les professionnels, il n'a jamais été imaginé de retranscrire une version réduite des données générées par les traitements du chef opérateur du son.

Il serait pertinent, à l'avenir, de travailler sur la réduction de la taille des données. Le travail consisterait en une saisie « plus grossière » des paramètres des métadonnées dynamiques. Par exemple, nous pourrions imaginer une modification de l'échelle de valeurs de la position des *faders*. En *Figure 3.14* ci-dessous, un exemple de modification d'échelle est présenté.

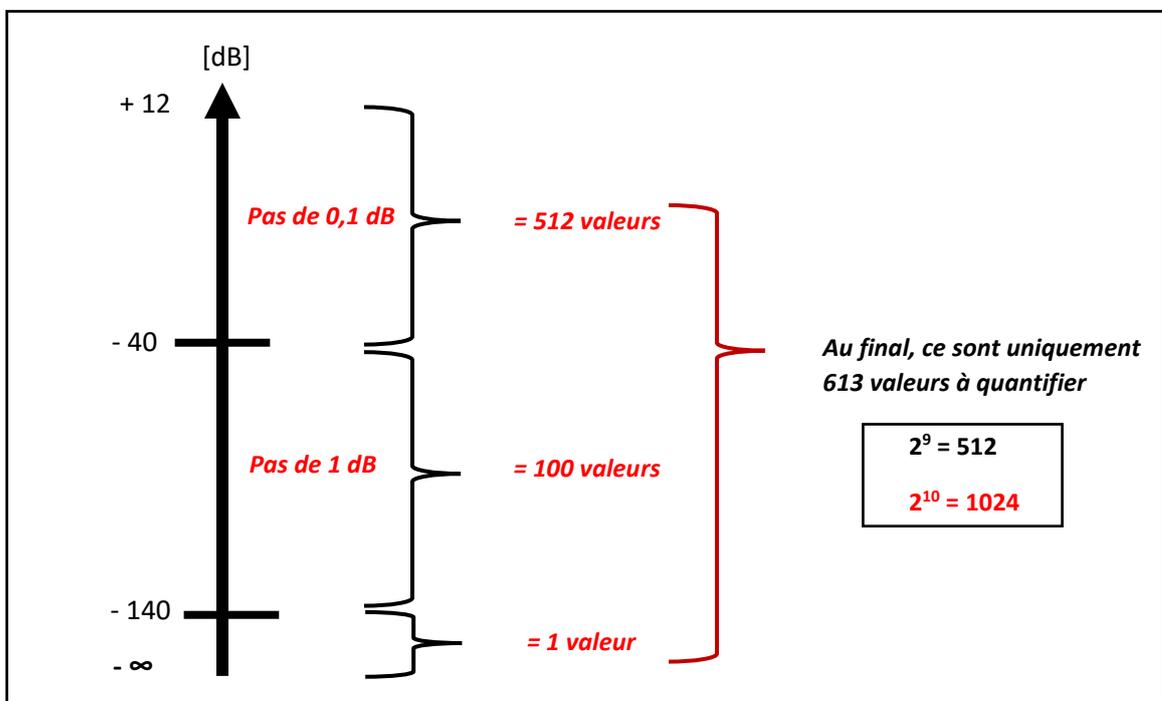


Figure 3.14 - Schéma présentant une possible évolution de l'échelle de gain

Dans l'exemple donné en *Figure 3.14*, utiliser 10 bits au lieu de 12 bits n'apparaît pas intéressant. Le gain en précision semble plus opportun que le gain de 2 bits par valeurs de position de *faders*. C'est un gain dérisoire en espace de stockage.

L'autre possibilité par rapport à cette problématique serait un travail à faire sur le lissage des points d'automation dès l'enregistrement comme le réalise déjà automatiquement les STAN.

III.2.3/ Enjeux liés à l'apparition des nouvelles métadonnées

L'implémentation de nouvelles métadonnées doit s'inscrire dans les méthodes de travail actuelles. Elles modifient les différentes pratiques et habitudes mises en place tout en employant les outils à disposition. Répondant à diverses problématiques, elles en soulèvent de nouvelles. À différents points stratégiques de la chaîne du son, les enjeux sont différents et de nouvelles interrogations se posent. Plus qu'une véritable avancée technologique qualitative, c'est bien une évolution des méthodes de post-production et une approche différente de la prise de son qui sont en jeu.

Néanmoins, depuis des décennies, des milliers de films sont réalisés sans avoir besoin de ces informations. La narration visuelle et sonore a-t-elle réellement besoin de ces évolutions technologiques pour mieux raconter des histoires ? Ces outils ne comptent pas bouleverser la fabrication d'un film mais tendent à en simplifier le processus.

Au moment du tournage, ces nouvelles métadonnées ne modifient en rien la pratique du chef opérateur du son. Elles sont pratiquement toutes générées automatiquement. Il paraît évident que les méthodes de travail employant une console de mixage analogique¹¹² ou encore une modulation à partir du préamplificateur de gain ne sont pas compatibles avec leurs utilisations. Ces métadonnées sont donc intimement liées à un emploi généralisé des surfaces de contrôle numérique. Se pose toutefois la question de l'interaction avec l'utilisateur. La question des différents paramètres accessibles à l'utilisateur sont présentés en *Partie III.3.2* lors de la présentation du démonstrateur.

Enfin, ces métadonnées n'affectent en rien l'emploi du *mixdown*. Son usage et sa forme restent inchangés. Tous les traitements dynamiques et fréquentiels sont appliqués directement au signal.

Métadonnées de filtrage

À l'étape du mixage et plus spécifiquement du prémixage paroles, les métadonnées de filtrage offrent une grande latitude aux mixeurs. Ce dernier peut s'appuyer sur le filtrage réalisé au moment du tournage, le modifier voire le supprimer en fonction des choix techniques et de mise en scène.

¹¹² À ce sujet, un échange entre les chefs opérateurs du son Stéphane Bucher, François Sempé et Thomas Lascar autour des consoles de mixage analogiques et des surfaces de contrôle est disponible à l'adresse URL suivante : <https://www.youtube.com/watch?v=x1Dj1d4T9NQ>. Cette vidéo a été réalisée en Juillet 2017 chez la société A4AUDIO.

Métadonnées de phase

Les métadonnées concernant la phase permettent au monteur paroles de savoir si le chef opérateur du son a réalisé ou non une opposition de phase. Cette information est importante lors de la remise en phase des pistes de microphones cravates avec la perche principale, étape importante et fastidieuse lors du montage paroles.

Métadonnées de panoramique

Lors du montage paroles, les métadonnées de panoramique permettent quant à elles d'appliquer rapidement aux pistes concernées, le gain réalisé en montage image sans devoir les écouter une à une pour être sûr des pistes mélangées dans le *mixdown*.

Métadonnées de modulation de volume

Les métadonnées liées à la modulation de volume entérinent le débat *pré* ou *post fader*. Désormais, la modulation de volume du chef opérateur du son est mise sous forme de points d'automatisation. Dès lors, il ne semble plus intéressant de travailler en *pré fader* car si la modulation venait à être modifiée, le monteur paroles aurait la possibilité de corriger la ligne d'automatisation créée au moment du tournage ou de repartir du signal non modulé, c'est-à-dire d'un signal en sortie du préamplificateur microphonique. Cette métadonnée permet donc de surmonter un hiatus historique.

Toutefois, cette idée de développement pourrait décrédibiliser le métier d'ingénieur du son et réduire la prise de son de plateau à une simple captation des différentes sources sonores où chacune de ces sources seraient enregistrées uniquement de manière « calibrée ». Finalement, ces métadonnées réduiraient le métier d'ingénieur du son à un « ramasseur de matières synchrones ». Si les enregistreurs actuels apportent des solutions qui engendrent des choix de moins en moins définitifs au moment de la prise de son, c'est laisser aux métiers de la post-production du son toutes les décisions. Au lieu de simplifier des étapes, ces métadonnées pourraient être perçues comme de nouvelles couches technologiques superficielles qui ne feraient que repousser les prises de décision.

Mon sentiment est qu'il existe une confusion majeure. C'est bien le chef opérateur du son qui prend toutes les décisions au moment de la prise de son mais transmet une matière sonore modulable en post-production. Les décisions sont prises à chaque étape et chaque professionnel s'appuie sur les décisions en amont tout en gardant une latitude maximale sur les données audio.

Un des enjeux important est l'intégration des métadonnées de modulation de volume lors de l'étape de conformation. Comment interpréter le gain appliqué par le monteur au moment du montage image ? Il ressort des discussions avec les monteurs paroles qu'il s'avère que les modifications de gain réalisées en montage image ne sont utilisées qu'à titre indicatif.

À court terme, ces nouvelles métadonnées ne peuvent pas s'inscrire dans un processus de conformation automatique. Il faudrait pour cela que leurs structures soient normalisées et compréhensibles par une STAN. C'est pourquoi, à l'aide du démonstrateur, nous envisagerons une autre manière de les lire dans une STAN.

III.3/ Implémentation de ces nouvelles métadonnées

Nous analyserons dans un premier temps, les implémentations envisageables au regard des outils disponibles sur le marché. Cette étude porte sur les vecteurs d'entrée disponibles dans les STAN actuelles. La première étape du travail est donc dépendante des solutions offertes par les fabricants. Suite à cet état des lieux et au regard des possibilités actuelles, nous proposerons un démonstrateur correspondant à la Partie Pratique du Mémoire (PPM). Le démonstrateur devra remplir les objectifs suivants :

- enregistrer les nouvelles métadonnées ;
- appliquer la mise en forme présentée en *Partie III.2.1* aux données enregistrées ;
- lire les nouvelles métadonnées dans une STAN.

III.3.1/ Présentation des vecteurs d'entrée

Ici, l'objectif est de définir l'outil permettant de lire les nouvelles métadonnées dans une STAN comme Pro Tools. Nous allons analyser l'ensemble des « portes d'entrée » disponibles actuellement. Les différentes possibilités offertes à l'utilisateur pour importer des données autres que le son sont multiples :

- | | |
|--------------------------------|-------------------|
| - protocole de communication ; | - fichiers WAVE ; |
| - fichier de session ; | - fichiers MIDI. |
| - super-conteneur audio ; | |

Les protocoles de communication tel que le Eucon et le HUI semblent des potentiels vecteurs dans l'importation de données d'automation dans une STAN. Or ces protocoles sont propriétaires et les spécifications sont indisponibles dans le domaine public.

Comme pour les protocoles de communication, les fichiers de session sont propriétaires et spécifiques à une STAN comme le fichier .ptx de Pro Tools employé depuis la version 10. Il est alors inenvisageable de travailler avec ce type de fichier et de produire un démonstrateur en se basant sur une technologie propriétaire.

Les super-conteneurs comme l'AAF, l'OMF et le XML apparaissent comme de parfaits vecteurs pour intégrer de nouvelles métadonnées. Sachant que l'AAF et l'OMF implémentent déjà les métadonnées de modulation de volume, nous avons tenté d'obtenir des informations dans les spécifications. Malgré le fait que la littérature soit importante pour le format AAF, les informations sur la modulation de volume ne sont pas disponibles. Il apparaît aussi difficile de travailler avec un outil où il n'est possible d'implémenter qu'un seul type de métadonnée (modulation de volume) alors que l'enjeu est d'implémenter les quatre métadonnées proposées dans le cadre de ce mémoire.

Le fichier WAVE ou encore BWF ont une structure de données normalisées. La possibilité d'enregistrer de nouvelles métadonnées dépend des fabricants d'enregistreurs alors que la possibilité de les lire est dépendante de la bonne volonté des STAN. C'est à travers ce constat qu'apparaît la difficulté de développer et d'implémenter de nouvelles métadonnées dans la chaîne du son. Les acteurs sont différents et ne possèdent pas les mêmes objectifs à court et moyen terme. Les enregistreurs numériques répondent à des demandes de preneurs de son, à des besoins provenant de la phase d'enregistrement du son, tandis que les STAN développent des outils dans des domaines différents comme la musique et le cinéma.

Le fichier MIDI est apparu comme la solution la plus viable dans l'exploitation d'un outil existant afin de véhiculer les métadonnées de modulation de volume dans une STAN. En employant le *Control Change n°7 (Channel Volume)*, retranscrire les modulations de volume apparaît envisageable. Or, le MIDI ne possède que 128 valeurs assignables alors que nous souhaitons en utiliser au moins 4096 (12 bits). De plus, le protocole ne permet pas d'obtenir de modulation de volume positive, c'est-à-dire un gain supérieur à 0. Malgré ces éléments, il semblait intéressant de voir si les métadonnées de volume étaient réellement importables dans une STAN avec un fichier MIDI. Nous avons donc créé un fichier MIDI à partir du langage Python en utilisant la bibliothèque *mido* présentée en *Figure 3.15*. Ici, deux points d'automation sont créés à deux positions temporelles différentes. Le test a été un échec

car malgré le fait de pouvoir importer ces deux points d'automation dans Pro Tools, il s'est avéré impossible de « copier-coller » la ligne d'automation de Volume MIDI de la piste MIDI sur une ligne d'automation de volume d'une piste audio.

```
from mido import Message, MidiFile, MidiTrack

mid = MidiFile()
track = MidiTrack()
mid.tracks.append(track)

track.append(Message('control_change', channel=0, control=7, value=85, time=0))
track.append(Message('control_change', channel=0, control=7, value=95, time=237))

mid.save('Test.mid')
```

Figure 3.15 - Présentation du code Python pour créer un fichier test MIDI

Au final, il s'avère inenvisageable d'employer les différents vecteurs actuels. Aucune de ces « portes » d'entrée ne permet d'inclure l'ensemble des nouvelles métadonnées proposées dans ce mémoire. Nous proposons alors de développer différents applicatifs sous le logiciel Max 7 de la société Cyclin'74 afin de créer différents outils nécessaires à l'implémentation de nouvelles métadonnées dynamiques dans une STAN.

III.3.2/ Phase d'enregistrement

Cette sous-partie présente le travail de programmation lié à la phase d'enregistrement des métadonnées. L'ensemble de la programmation s'est cantonné à deux logiciels : les logiciels Max 7 et Arduino.

Comme présentés en début de *Partie III.3*, nous proposons de remplir les deux premiers objectifs :

- enregistrer les nouvelles métadonnées ;
- appliquer la mise en forme présentée en *Partie III.2.1* aux métadonnées enregistrées.

Développé à l'origine par Miller Puckette à l'IRCAM, Max 7 correspond à un logiciel de programmation informatique visuelle permettant aux musiciens, performeurs, chercheurs et développeurs de créer graphiquement des programmes sans écrire la moindre ligne de code. Il permet de réaliser divers outils comme du traitement du signal, de l'enregistrement ainsi que du contrôle d'instrument MIDI.

La particularité de ce logiciel est d'avoir une communauté d'internautes très puissante fournissant de nombreux exemples et des bibliothèques externes supplémentaires.

Max utilise la notion de *patcher* dont l'extension de fichier est *.maxpat*. Le *patcher* est une fenêtre où des boîtes communiquent entre elles via des fils. Chacune de ces boîtes possède une fonction bien spécifique (objet, message, etc.). À travers le chemin du signal réalisé par l'utilisateur, c'est-à-dire les fils reliant les boîtes les unes aux autres, les boîtes ont la possibilité de recevoir un message, d'appliquer une fonction au message reçu et de renvoyer une information en sortie. Enfin, l'utilisateur dispose de plusieurs couches de programmation, en utilisant des « sous-*patcher* », des *patchers* dans un *patcher*.

En plus de la programmation réalisée sous Max 7, nous décidons d'utiliser deux potentiomètres à glissière. Au regard des solutions actuelles pour communiquer avec Max, nous choisissons d'utiliser une carte *Arduino* « Uno » (cf. *Annexe L* présentant les différentes caractéristiques de la carte). Ce modèle de microcontrôleur¹¹³ possède les avantages suivants :

- nombre important d'entrées et sorties analogiques et numériques ;
- alimentation simple par un port USB - tension de 5 Volts (V) ;
- communauté d'utilisateurs importante.

Pour configurer la carte *Arduino*, nous transférons la bibliothèque *ConfigurableFirmata*¹¹⁴. Cette bibliothèque nous permet d'établir la communication avec le logiciel Max. Enfin, nous utilisons un *patcher* de Max 7 préexistant nommé *Maxuino*¹¹⁵, *open source* et développé de manière collaborative. C'est à travers ce *patcher* que nous recevons la tension aux bornes du potentiomètre à glissière connecté à l'entrée analogique n°0 de la carte *Arduino*. À partir du signal délivré par la carte *Arduino*, nous réalisons un changement d'échelle pour transformer cette tension en un gain numérique allant de -140 dB à + 12 dB.

¹¹³ Un *microcontrôleur* correspond à un circuit intégré rassemblant les éléments essentiels d'un ordinateur : processeur, mémoire (morte et vive), unités périphériques et interfaces d'entrées-sorties. De petite taille, il est destiné à un applicatif spécifique.

¹¹⁴ Disponible gratuitement à l'adresse URL : <https://www.arduinolibraries.info/libraries/configurable-firmata>

¹¹⁵ Disponible gratuitement à l'adresse URL : <http://www.maxuino.org/>

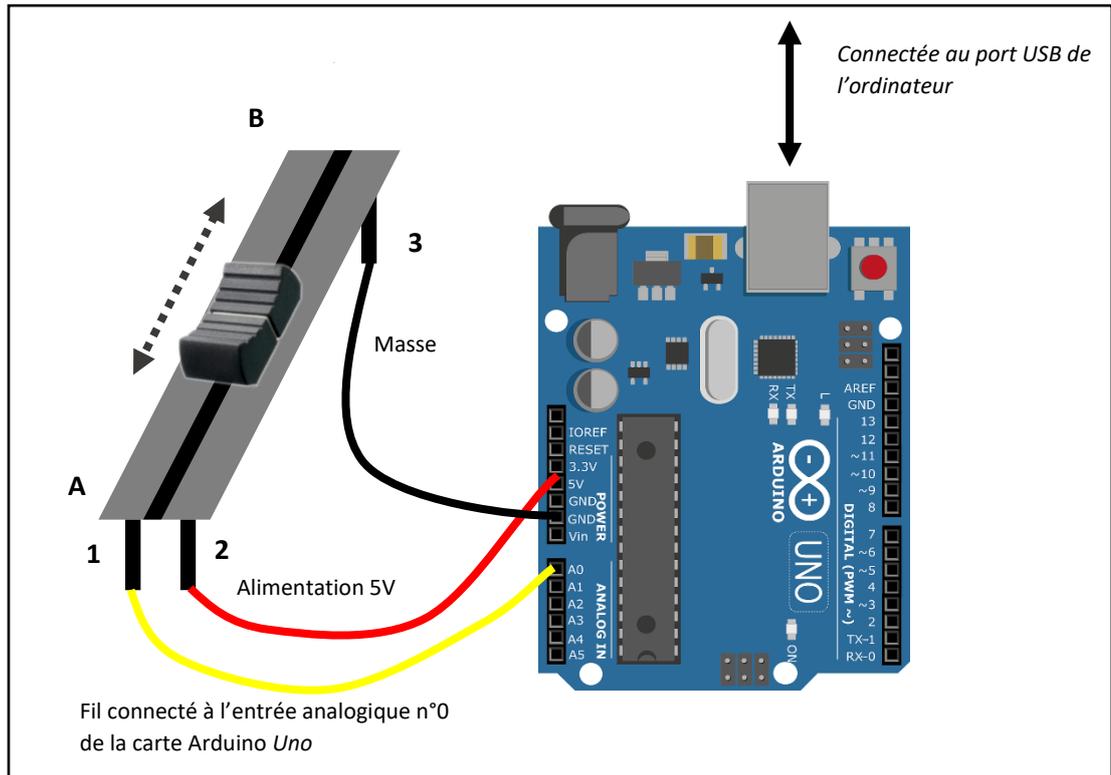


Figure 3.16 - Présentation des branchements entre le potentiomètre et la carte Arduino

D’après la documentation de la carte, la tension de l’entrée n°0 de la carte Arduino *Uno* est convertie en numérique sur 10 bits (1024 valeurs). Ainsi, lorsque le *fader* est en bas (position A de la Figure 3.16 ci-dessus), la tension en sortie du potentiomètre est nulle et sa valeur numérique est 0 tandis que lorsque le *fader* est en haut (position B), la tension en sortie du potentiomètre est de 5 V et sa valeur numérique est égale à 1023.

Dans le cadre du développement du projet sous Max 7, l’objectif principal est d’enregistrer en parallèle du flux audio différents paramètres codés en XML. Il est décidé dès le départ d’enregistrer deux fichiers. Le premier est un fichier WAVE contenant uniquement les données audio tandis que le second est un fichier au format XML contenant les nouvelles métadonnées.

La première question qui s’est posée lors du développement du projet est « quelle interaction l’utilisateur possède-t-il avec ces nouvelles métadonnées ? ». Ce questionnement est intrinsèquement lié à l’interface graphique du *patcher* (appelée aussi *Graphic User Interface* (GUI) en anglais), plateforme dont l’intuitivité est au cœur des préoccupations. En sachant que l’un des premiers éléments qui détermine l’efficacité ou non d’une interface est le nombre de « clics » demandés à l’utilisateur afin d’accéder à sa requête. Chaque bouton de l’interface développée est le plus « transparent » possible et s’insère dans des méthodes déjà employées pour utiliser les traitements

dynamiques, fréquentiels, de phase et de panoramique. Voici en *Figure 3.17* ci-dessous, l'interface de contrôle du démonstrateur.

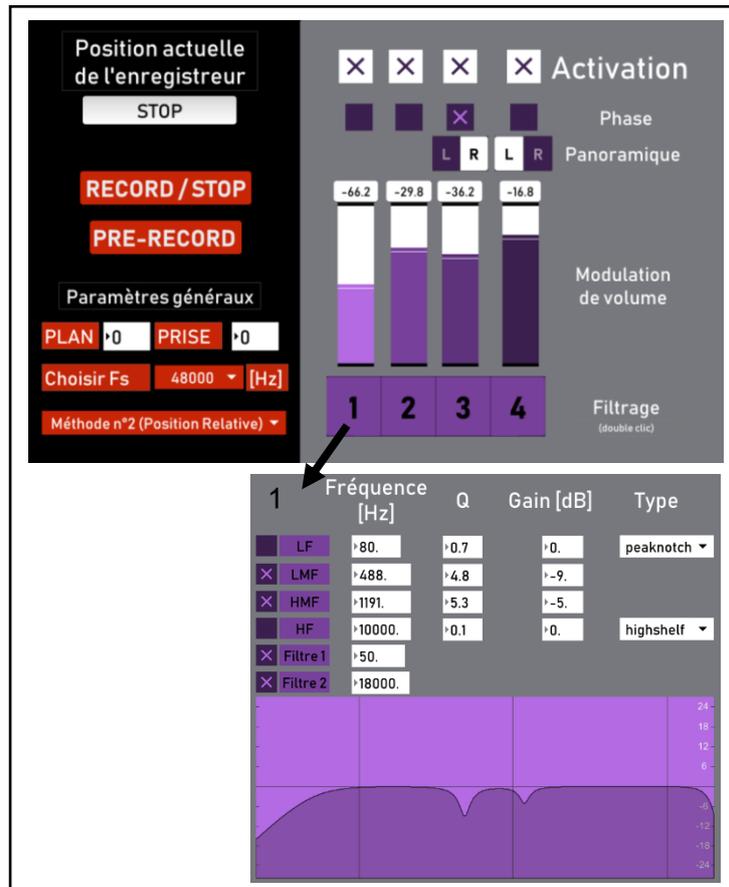


Figure 3.17 - Présentation de l'interface de contrôle dans Max 7

Les paramètres situés à gauche (dans la partie sur fond noir) sont de deux natures. En haut, ils correspondent au transport (RECORD/STOP et PRE-RECORD). Afin d'écouter les entrées de la carte son interne ou l'interface audio sans toutefois enregistrer le moindre fichier, l'utilisateur utilise le bouton PRE-RECORD. En appuyant une première fois sur le bouton RECORD/STOP, un fichier WAVE est créé et enregistre le signal audio. Durant cette période d'enregistrement, l'utilisateur peut modifier les paramètres de traitement. En appuyant une seconde fois sur le bouton RECORD/STOP, le fichier WAVE est créé et une requête est envoyée afin de débiter la création du fichier XML.

En-dessous, se situent les « paramètres généraux » qui permettent de définir :

- nom des fichiers WAVE et XML ;

Avant de débiter l'enregistrement, l'utilisateur peut modifier le nom du fichier WAVE et XML respectivement sous la forme *scene00take00.wav* et *scene00take00.xml*.

- fréquence d'échantillonnage du fichier (Fs) ;
- choix de la méthode d'enregistrement des métadonnées de modulation de volume.

L'utilisateur décide entre une des deux méthodes d'enregistrement des métadonnées de modulation de volume (échantillonnage des *faders* avec FsF ou affectation d'une valeur temporelle *relative*).

Dans la partie droite (fond gris), l'utilisateur retrouve l'ensemble des boutons qui contrôlent les différents paramètres de traitement. Il peut, à tout moment, modifier le filtrage, le volume, le panoramique et la phase de chacune des pistes. Ce sont au final les états de ces boutons au moment de l'enregistrement qui sont mis sous forme de métadonnées.

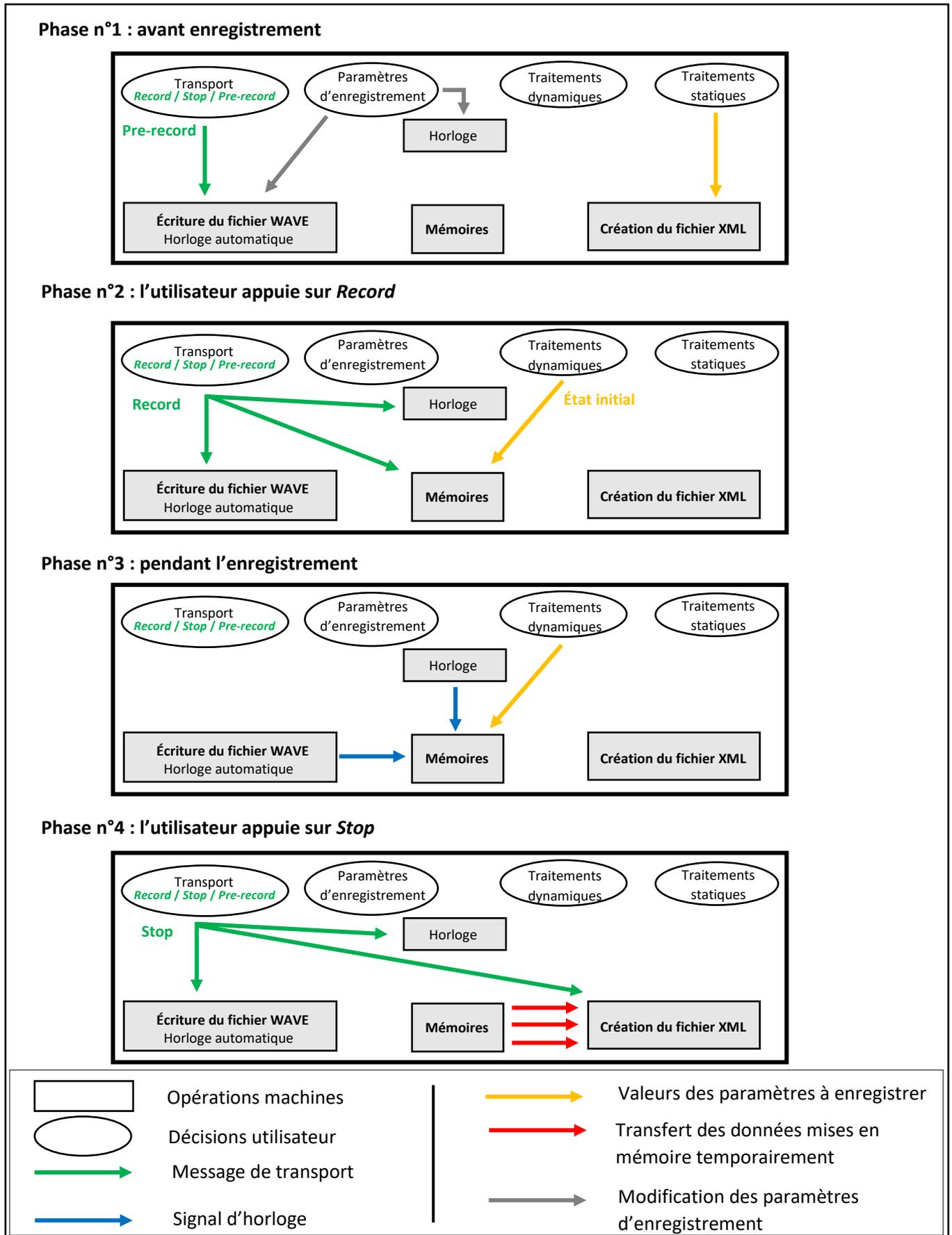


Figure 3.18 - Synoptique du fonctionnement des quatre étapes de la partie enregistrement du démonstrateur
99/138

Au niveau de la programmation dans Max, nous présentons les différents outils utilisés à chacune des trois étapes. Voici l'ensemble des *patchers* développés et regroupés dans le *patcher* principal :

- *patcher* n°1 : Fabrication de l'horloge n°1 ;
- *patcher* n°2 : Communication avec la carte son interne ou interface audio externe ;
- *patcher* n°3 : Fabrication du nom de fichier ;
- *patcher* n°4 : Mémoire temporaire des données de modulation de volume (méthode n°1) ;
- *patcher* n°5 : Mémoire temporaire des données de modulation de volume (méthode n°2) ;
- *patcher* n°6 : Mémoire temporaire des données de filtrage ;
- *patcher* n°7 : Mémoire temporaire des données de phase ;
- *patcher* n°8 : Mémoire temporaire des données du panoramique ;
- *patcher* n°9 : Fabrication du fichier XML.

Voici le fonctionnement général des quatre étapes présentées en *Figure 3.18* :

- Avant l'enregistrement

Avant de débiter l'enregistrement, l'utilisateur renseigne les paramètres généraux. En modifiant le numéro de scène et de prise, il déclare différentes instances dans le *patcher n°3* qui fabrique le nom du fichier mis en mémoire temporairement. L'utilisateur peut aussi définir la méthode d'enregistrement des métadonnées de modulation de volume.

Si l'utilisateur décide d'écouter les signaux en entrée de la carte son ou de l'interface audio, alors il appuie sur le bouton Pre-record qui communique directement avec le *Patcher n°2*.

Enfin, il peut modifier les paramètres de traitement avant de débiter l'enregistrement. Les paramètres de filtrage et de panoramique sont enregistrés temporairement et respectivement dans les *Patchers n°6* et *8*. Les données sont enregistrées temporairement dans des boîtes dites *message* (). En *Figure 3.19*, un exemple de ce type de boîte est présenté.

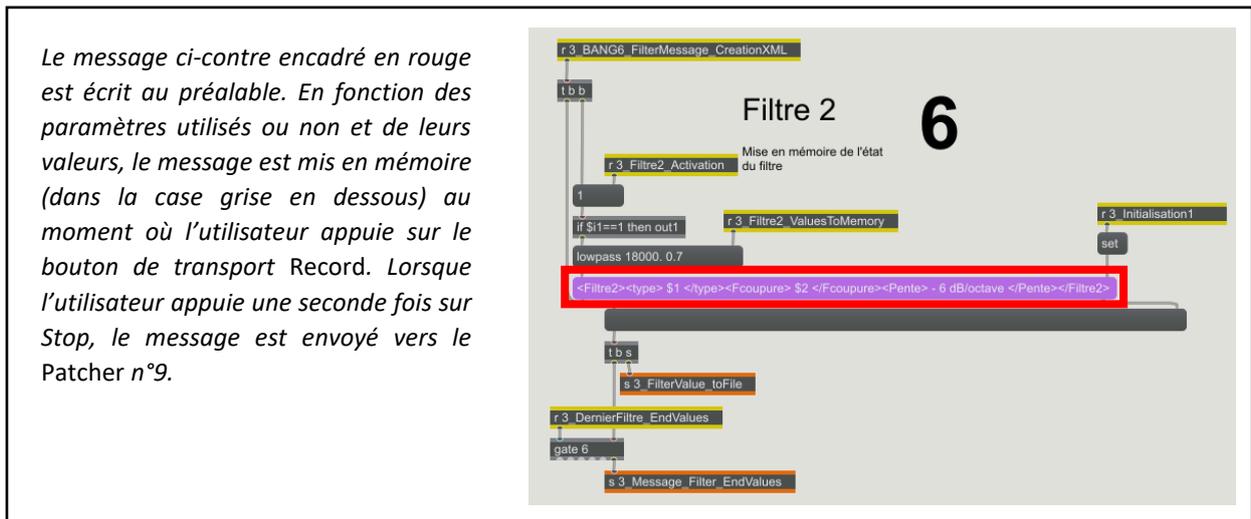


Figure 3.19 - Présentation de l'enregistrement temporaire de la métadonnée Filtrage dans Max 7

- L'utilisateur appuie sur *Record*

En appuyant sur *Record*, différents instances s'enclenchent. La première active la carte son dans le *patcher n°2* et efface l'ensemble des informations stockées dans les *patchers n°4, 5, 6, 7 et 8*. La seconde instance active l'horloge du *patcher n°1* afin de l'envoyer au *patcher n°4*. L'état initial de tous les paramètres de traitement sont envoyés aux *patchers n°4, 5, 6, 7 et 8*.

- Au cours de l'enregistrement

L'utilisateur modifie l'ensemble des traitements dynamiques (modulation de volume et phase) et les différents états des paramètres sont continuellement enregistrés dans les *patchers n°4 ou 5* (en fonction de la méthode choisie) pour la modulation de volume et n°7 pour la phase.

Ces trois *patchers* utilisent la boîte `coll`. Elle permet d'enregistrer des données en temps réel dans un tableau où le numéro de ligne est incrémenté automatiquement.

Le *patcher n°9* utilise comme boîte centrale, `sadam.rapidXML`¹¹⁶. Cette boîte a été créée par Adam Siska et permet de lire, écrire et enregistrer un fichier XML. Au niveau du fonctionnement de ce *patcher*, l'ensemble des actions sont automatiques et déclenchées par la machine. Son fonctionnement est basé sur plusieurs types de boîtes :

- message préinscrit ;

C'est le cas pour les balises qui structurent le fichier XML.

```
<?xml version="1.0" encoding="UTF-8"?><Traitements>
```

Première boîte à être lue dans le Patcher n°9

- message créé avant l'enregistrement ;

Ces messages concernent l'ensemble des métadonnées statiques (panoramique, filtrage) et les différents paramètres généraux.

- messages arrivant en cascade.

Ces messages concernent les paramètres des métadonnées dynamiques (modulation de volume et phase).

III.3.3/ Phase de lecture

Dans cette partie, nous présentons le travail de programmation lié à la phase de lecture des métadonnées du démonstrateur. L'ensemble de la programmation a été réalisée sous Max 7. Cette partie du démonstrateur propose de remplir le dernier objectif : lire les nouvelles métadonnées dans une STAN.

Ce module de restitution des nouvelles des métadonnées répond à différents objectifs :

- interaction simple avec l'utilisateur ;
- lecture et extraction des données d'un fichier XML ;
- communication avec une STAN.

¹¹⁶ Disponible à l'adresse URL : http://www.maxobjects.com/?v=libraries&id_library=154

Voici le fonctionnement du démonstrateur présenté en *Figure 3.21* ci-dessous.

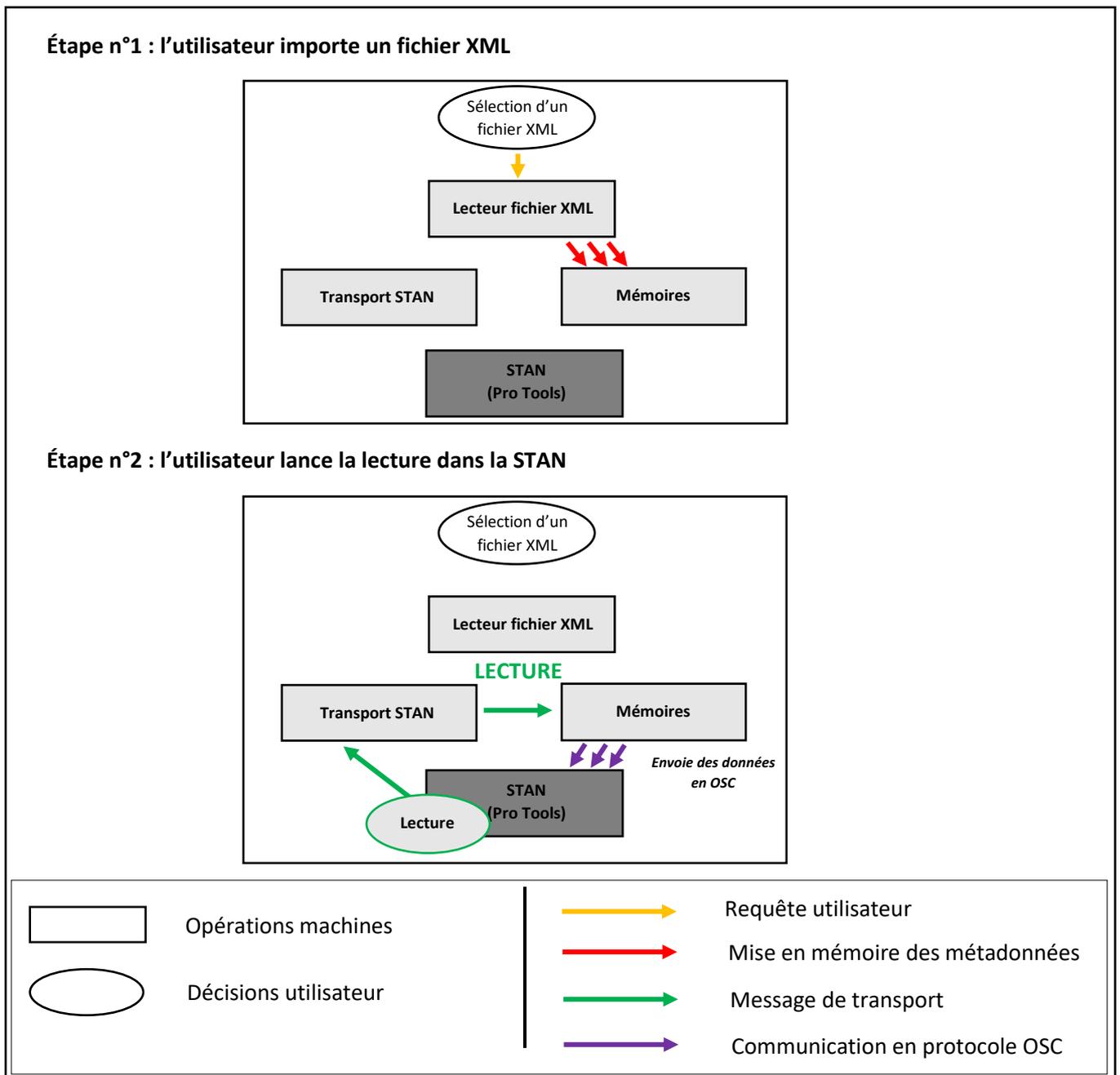


Figure 3.21 - Schéma fonctionnel du module de lecture du démonstrateur

Cette partie du démonstrateur se compose de trois *patcher* développés et regroupés dans un *patcher* principal :

- *patcher* n°1 : Lecture et extraction des données XML ;
- *patcher* n°2 : Mémoire temporaire ;
- *patcher* n°3 : Communication avec la STAN ;

Voici le fonctionnement général dans les deux étapes de cette phase de lecture :

- Ouverture du Fichier XML

L'utilisateur appuie sur le bouton *RECHERCHER* qui lance la lecture du fichier XML. Le *patcher* n°1 utilise de nouveau la boîte `sadam.rapidXML`. Cette dernière permet de lire l'ensemble du contenu des balises XML et envoie les différentes informations au *patcher* n°2. Les données sont alors stockées temporairement dans les boîtes `coll`. Au final, nous utilisons les mêmes outils que pour la première partie du démonstrateur.

- Lecture dans la STAN

Au moment où l'utilisateur lance la lecture dans la STAN, nous recevons par la boîte `hostsync~` les informations de transport de la STAN dans Max. Cette boîte permet de récupérer différentes informations provenant d'une STAN en utilisant le protocole *ReWire*¹¹⁷. Lorsque l'information de « lecture » est reçue dans Max 7, différentes instances sont enclenchées.

La première active l'extraction des données du *patcher* n°2. Ces données sont alors d'adressées au *patcher* n°3 qui permet d'envoyer en *Open Sound Control*¹¹⁸ (OSC) les différents paramètres. Nous utilisons alors la boîte `udpsend localhost 4001` qui transmet avec succès les différents paramètres vers la STAN où se trouve le *plug-in* *Tosca*¹¹⁹ en insert d'une des pistes. Voici les lignes d'automation générées dans la STAN Pro Tools en *Figure 3.22* à partir du fichier présenté en *Annexe M*. On observe qu'un lissage des données a été réalisé automatiquement.

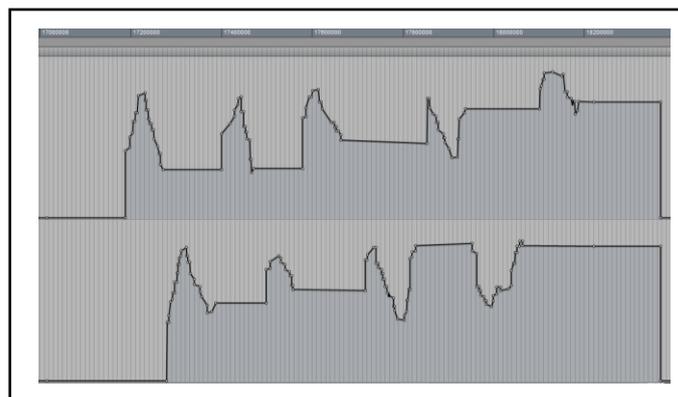


Figure 3.22 - Courbes d'automation générées dans Pro Tools à partir du module de lecture

¹¹⁷ Créé en 1998 par les sociétés Steinberg, Propellerhead Software et Rebirth, le *ReWire* correspond à un protocole d'échange d'informations entre plusieurs STAN. Par exemple, il permet de synchroniser la lecture de deux STAN différentes.

¹¹⁸ L'*Open Sound Control* (OSC) est un format de transmission de données entre ordinateurs conçu pour un fonctionnement « temps réel ». Il utilise le réseau pour échanger les données.

¹¹⁹ *Tosca* est l'acronyme de Thibaut's OpenSoundControl Automation. Il correspond à un *plug-in* au format AU, VST, VST3 et AAX qui n'effectue aucune traitement audio. Son fonctionnement permet d'envoyer et de recevoir des paramètres d'automation via le réseau.

L'utilisateur n'a plus qu'à copier et coller les lignes d'automatisation sur les pistes des clips correspondants. À travers le détournement de l'outil *ToscA*, nous avons pu intégrer dans Pro Tools les métadonnées dynamiques de modulation de *fader*. Sachant que cette manipulation est alors possible pour tous les paramètres de modulation de volume, il est tout à fait envisageable de lire d'autres métadonnées dynamiques telles que les métadonnées de phase.

Au cours de ce dernier chapitre, nous avons pu établir les différents paramètres caractérisant les traitements que nous souhaitons mettre sous forme de métadonnées. En établissant une mise en forme au format XML, nous avons soulevé de nouvelles problématiques comme le poids de ces données.

Le démonstrateur propose de reprendre l'ensemble des éléments définis en début de chapitre et de les employer dans une STAN. Finalement, il remplit les deux objectifs fixés en début de mémoire, il permet d'enregistrer et de lire de nouvelles métadonnées dans une STAN.

Conclusion

L'objectif de ce mémoire est de proposer une solution répondant à des enjeux réels de terrain dans l'utilisation de fichiers audio dans une chaîne du son au cinéma. Une analyse des métadonnées professionnelles existantes dans le domaine public a permis de mettre en lumière un premier constat : l'impossibilité pour les métiers de la post-production du son de connaître les différents traitements réalisés par le chef opérateur du son en direct au moment du tournage. Pourtant, en étudiant les pratiques employées en prise de son, de réels enjeux tant techniques qu'artistiques en découlent. C'est bien la voix, source sonore incontournable dans une bande son, qui est souvent impliquée dans ces traitements.

Les traitements de modulation de volume, de phase, de panoramique et de filtrage sont des éléments bien définis dans la pratique des professionnels. Les métadonnées que nous proposons et qui s'y réfèrent répondent à des problématiques longuement évoquées au sein de différents ateliers ou conférences d'associations de professionnels du son comme l'Association Française du Son à l'Image (ASFI).

Au cours du *Chapitre III*, nous avons pu caractériser l'ensemble des paramètres qui définissent ces métadonnées. À partir de cette approche tant théorique que fondamentale, ce mémoire établit une mise en forme pour ces nouvelles métadonnées. En s'appuyant sur ces éléments, nous présentons un démonstrateur afin d'enregistrer et de lire les nouvelles métadonnées dans une Station de Travail Audionumérique (STAN).

Au regard des résultats obtenus, le démonstrateur répond aux deux objectifs fixés au début du mémoire : enregistrer et traiter de nouvelles métadonnées. Il répond à une problématique de faisabilité. Cette étude pose donc la « première » pierre d'un développement futur de nouvelles métadonnées mais surtout, elle propose une nouvelle manière d'enregistrer les traitements réalisés par le chef opérateur du son en employant des méthodes dites virtuelles utilisées dans une STAN. Les traitements sont désormais non-destructifs et conservés sous forme de lignes d'automation.

La démarche de ce mémoire n'est pas d'établir une norme mais au contraire d'offrir de nouvelles possibilités dans l'intégration de métadonnées dynamiques. À travers cette recherche, nous avons vu que cette étude se trouvait à la rencontre de nombreux domaines, de l'étape du tournage aux métiers de la post-production.

Bibliographie

Ouvrages

BAILBLE, Claude (sous la dir. de), *Le son au cinéma*, Paris, Ministère de l'Éducation et de la Culture, coll. Langage et techniques, 1993.

BALIBAR, Lucien, *La chaîne du son au cinéma et à la télévision : De la prise de son à la post-production*, Paris, Dunod, coll. Audio-Photo-Vidéo, 2015.

CHION, Michel, *La Voix au Cinéma*, Paris, Cahiers du Cinéma, coll. Essais, 1982, p. 18 - 20.

LEROUGE, Claude, *Sur les 100 années, le cinéma sonore*, Paris, Dujarric, 1996.

MAGNIER, Vincent, *Guide de la prise de son pour l'image – Reportage, documentaire, fiction en radio et télévision 2^e édition*, Paris, Dunod, 2007, réed. coll. Audio-Photo-Vidéo, 2015, p.64 – 65.

MERCIER, Denis (sous la dir. de), *Le livre des techniques du son - Tome 2 : La technologie - 4^{ème} édition*, Paris, Fréquences, 1988, réed. Paris, Dunod, coll. Audio-Photo-Vidéo, 2012, p. 344 - 346.

NOUGARET, Claudine, CHIABAUT, Sophie, *Le son direct au cinéma*, Paris, FEMIS – Institut de Formation et d'enseignement pour les métiers de l'image et du son, 1997.

PURCELL, John, *Dialogue Editing for Motion Pictures : A Guide to the Invisible Art*, Burlington, Focal Press, 2013, pages 76 - 80.

Articles de périodique

BARNIER, Martin, « Les premiers ingénieurs du son français », 1895. *Mille huit cent quatre-vingt-quinze*, Histoire des métiers du cinéma en France avant 1945, n°65, 2011.

CANONVILLE, Christian, « De quelques implications esthétiques des innovations techniques dans la dimension sonore des films », *Cahier Louis-Lumière* n°1, automne 2003, p. 92 - 107.

ISAACS, Paul, « iXML », Line UP, *Line Up (Journal of the Institute of Broadcast Sound)* n°109, février-mars 2007, p.20 – 22.

Mémoires et thèses

CHARLEMAGNE, Paul, *Psychoacoustique : Mesure subjective des paramètres d'intensité, de fréquence et de temps*, mémoire sous la direction de Yves LASRY, Université Henri Poincaré – Nancy Université, Faculté de Pharmacie, 2010.

DEBAECKER, Jean, *De l'usage des métadonnées dans l'objet sonore*, Thèse sous la direction de Widad Mustafa El Hadi, Sciences de l'informatique et de la Communication, Université Charles-de-Gaulle – Lille 3, École doctorale S.H.S. - E.D. 473 - Laboratoire Geriico, E.A. 4073, 2012.

DELPLANCQ, Jeanne, *La modélisation des connaissances dans le montage son au cinéma*, mémoire sous la direction de Frédéric CHANGENET et Jean-Pierre HALBWACHS, École Nationale Supérieure Louis Lumière, section son, 2009.

DURAND, Hugo, *Etude et perspectives sur la recherche de données en sonothèque, en vue d'une implantation collaborative à l'École Nationale Supérieure Louis-Lumière*, mémoire sous la direction de Sylvain LAMBINET et Eric LONNI, École Nationale Supérieure Louis Lumière, section son, 2017.

GUILLAUME, Olivier, *L'organisation en postproduction sonore cinématographique*, mémoire sous la direction de Jean-Pierre HALBWACHS et Selim AZZAZI, École Nationale Supérieure Louis Lumière, section son, 2010, p. 59 - 62.

LAGARENNE, Nicolas, ORANTIN, Jonas, *Le recalage temporel des signaux issus des microphones perche et HF par le monte-parole au cinéma*, mémoire sous la direction de Claude GAZEAU, Matthieu FICHET, Laurent MILLOT et Vincent MONTROBERT, École Nationale Supérieure Louis Lumière, section son, 2015.

MESANGE, Baptiste, *MUDO : Dimensionnement et implémentation d'un enregistreur portable modulaire*, mémoire sous la direction de Sylvain LAMBINET et Mattéo CASSONNET, École Nationale Supérieure Louis Lumière, section son, 2017.

RACINE, Dorian, *Les microphones numériques et la prise de son à l'image*, mémoire sous la direction de Frédéric BELIN et Axel BRISARD, Université d'Aix-Marseille, Département Sciences, Arts et Techniques de l'Image et du son, 2016.

VUILLET, Romain, *De quelques nouvelles potentialités en postproduction cinématographique apportées par l'usage de l'enregistreur multipiste*, mémoire sous la direction de Daniel SOBRINO et Christian CANONVILLE, École Nationale Supérieure Louis Lumière, section son, 2006.

Sources Internet

A4AUDIO, « Podcast : Débat sur l'intelligibilité », Youtube, octobre 2017, URL : https://www.youtube.com/watch?v=lhSvFRrPS_k

A4AUDIO, « Podcast : Console ou surface de contrôle ? », Youtube, juillet 2017, URL : <https://www.youtube.com/watch?v=x1Dj1d4T9NQ>

ADLER, Laure, « Emission HORS-CHAMPS – Jean-Pierre LAFORCE », France Culture, mai 2013, URL : <https://www.franceculture.fr/emissions/hors-champs/jean-pierre-laforce>

ANDREWS, John, « AES-X155 Initiation », Audio Engineering Society, mai 2008, URL : <http://www.aes.org/standards/meetings/init-projects/aes-x155-init.cfm>

Association Française du Son à l'Image [forum de discussion], « Montage des Dialogues d'après Multipiste », 2008, URL : <http://archive.afsi.eu/forum/montage-des-dialogues-d-apr%E2%88%9A%C2%AEs-multipiste>

Association Française du Son à l'Image [forum de discussion], « Montage des directs...Ou démontage du direct ? », 2008, URL : <http://archive.afsi.eu/forum/montage-des-directsou-d%E2%88%9A%C2%A9montage-du-direct>

Association Française du Son à l'Image [forum de discussion], « Mixdown du direct », 2012, URL : <http://archive.afsi.eu/node/4223>

Association Française du Son à l'Image [forum de discussion], « Montage de directs et grosse fatigue », 2010, URL : <http://archive.afsi.eu/forum/montage-de-directs-et-grosse-fatigue>

Commission Supérieure Technique de l'image et du son, « Les rencontres de la CST », CST, juin 2017, URL : <https://www.cst.fr/la-cst/les-rencontres-de-la-cst/>

FERENCZI, Aurélien, « Il y a des sons qui se perdent... », Télérama, décembre 2009, URL : <http://www.telerama.fr/cinema/il-y-a-des-sons-qui-se-perdent,50545.php>

Gallery UK, « iXML Specification Revision - Standard for embedded metadata in production media files », Site officiel du iXML, mars 2018, URL : <http://ixml.info/>

GROULT, François, « Compte rendu - Traitement des dialogues en auditorium », Association Française du Son à l'Image, décembre 2008, URL : <https://www.afsi.eu/articles/22581-compte-rendu-traitement-des-dialogues-en-auditorium>

LEBON, Thierry, « Compte-rendu - L'atelier « Le son direct » », Association Française du Son à l'Image, juin 2013, URL : <https://www.afsi.eu/articles/15233-atelier-le-son-direct>

LONNI, Éric, « Compte-rendu - L'atelier « Que se cache-t-il derrière les fichiers wave et les metadatas » », Association Française du Son à l'Image, septembre 2014, URL : <http://archive.afsi.eu/node/5271?tab2=3>

MAGINIER, Vincent, « L'évolution de la prise de son en tournage audiovisuel », Ina, janvier 2013, URL : <https://www.ina-expert.com/e-dossier-de-l-audiovisuel-le-son-dans-tous-ses-etats/l-evolution-de-la-prise-de-son-en-tournage-audiovisuel.html>

MAGNIER, Vincent, « Compte-rendu « Exploiter le son direct du plateau à l'auditorium avec le nouveau Aaton Cantar X3 », Association Française du Son à l'Image, février 2016, URL : <https://www.afsi.eu/articles/24659-le-nouveau-aaton-cantar-x3>

MORANT (de), François, « Entretien avec Pierre Gamet », Association Française du Son à l'Image, février 2012, URL : <https://www.afsi.eu/articles/12770-entretien-avec-pierre-gamet>

PECCATTE, Patrick, « Métadonnées : une initiation - Dublin Core, IPTC, Exif, RDF, XMP, etc. », Soft Experience, PIM – Philosophie, Informatique, Mathématiques, décembre 2007, URL : <http://peccatte.karefil.com/Software/Metadata.htm>

SSEI [pseudonyme], « MACKIE HUI MIDI protocol - The results of a 2-day reverse-engineering-session by theageman », REAPER [Forum], janvier 2010, URL : <https://forum.cockos.com/showthread.php?t=101328>

Spécifications scientifiques

Advanced Media Workflow Association, Spécification Advanced Authoring Format, Advanced Authoring Format – Preliminary Draft, Version 1.1, 1999.

Advanced Media Workflow Association, Spécification Advanced Authoring Format, Advanced Authoring Format (AAF) – Edit Protocol, avril 2008.

Audio Engineering Society, Article de la 121^{ème} convention de l'AES à San Francisco, EuCon : An Object-Oriented Protocol for Connecting Control Surfaces to Software Applications, octobre 2006.

Apple, Spécification Audio Interchange File Format, Audio Interchange File Format : « AIFF » - A standard for Sampled Sound Files, Version 1.3, janvier 1989.

Extensible Markup Language (XML), Recommandation W3C, Extensible Markup Language (XML) 1.1 (Second Edition), Version 1.1, septembre 2006, URL : <https://www.w3.org/TR/2006/REC-xml11-20060816/>

IRCAM, CARPENTIER, Thibaut, Article des Journées d'Informatique Musicale (JIM), TosCA : un plugin de communication OSC pour le mixage spatialisé orienté objet, 2015.

Microsoft, Spécification Waveform Audio File Format, New Multimedia Data Types and Data Techniques, Révision 3.0, avril 1994.

MIDI Manufacturers Association, Spécification Musical Instrument Digital Interface, General MIDI 2, Version 1.2a, février 2007.

Union Européenne de Radio-Télévision, Recommandation EBU R 099, « Unique » Source Identifier (USID) for use in the <OriginatorReference> field of the Broadcast Wave Format, octobre 2011.

Union Européenne de Radio-Télévision, Spécification EBU-TECH 3285, Specification of the Broadcast Wave Format (BWF) – A format for audio data files in broadcasting, Version 2.0, mai 2011.

Union Européenne de Radio-Télévision, Supplément 1 EBU-TECH 3285, BWF – A format for audio data files in broadcasting – Supplement 1 – MPEG audio, juillet 1997.

Union Européenne de Radio-Télévision, Supplément 2 EBU-TECH 3285, BWF – A format for audio data files in broadcasting – Supplement 2 – Capturing Report, juillet 2001.

Union Européenne de Radio-Télévision, Supplément 3 EBU-TECH 3285, BWF – A format for audio data files in broadcasting – Supplement 3 – Peak Envelope Chunk, juillet 2001.

Union Européenne de Radio-Télévision, Supplément 4 EBU-TECH 3285, BWF – A format for audio data files in broadcasting – Supplement 4 – <link> Chunk, avril 2003.

Union Européenne de Radio-Télévision, Supplément 5 EBU-TECH 3285, BWF – A format for audio data files in broadcasting – Supplement 5 – <axml> Chunk, juillet 2003.

Union Européenne de Radio-Télévision, Supplément 5 EBU-TECH 3285, BWF – A format for audio data files in broadcasting – Supplement 6 – Dolby Metadata, <dbmd> chunk, octobre 2009.

Union Européenne de Radio-Télévision, Spécification EBU-TECH 3306, MBWF / RF64 : An extended File Format for Audio – A BWF-compatible multichannel file format enabling file sizes to exceed 4 Gbyte, juillet 2009.

Union Européenne de Radio-Télévision, Spécification EBU-TECH 3293, EBU CORE METADATA SET (EBUCore), Version 1.8, octobre 2017.

Union Européenne de Radio-Télévision, Spécification EBU-TECH 3295, P_META – Metadata Library, Version 2.2, septembre 2011.

Union Européenne de Radio-Télévision, **DELVIN, Bruce**, Révision technique, MXF – the Material eXchange Format, juillet 2002.

Union Européenne de Radio-Télévision, **GILMER, Brad**, Révision technique, AAF – the Advanced Authoring Format, juillet 2002.

Table des illustrations

Chapitre I

Figure 1.1 - Représentation simplifiée du fichier WAVE	12
Figure 1.2 - Exemple d'entrelacement d'un fichier polyphonique contenant 4 canaux.....	14
Figure 1.3 - Représentation des tranches historiques du fichier WAVE	15
Figure 1.4 - Structure générale du format Broadcast Wave	17
Figure 1.5 - Tableau présentant les différents types de métadonnées professionnelles en audio	19
Figure 1.6 - Tableau des métadonnées appartenant de la tranche BROADCAST EXTENSION.....	20
Figure 1.7 - Tableau des différentes balises dans la tranche iXML	26
Figure 1.8 - Capture d'écran du Cantar X-3 (version 2.23).....	35
Figure 1.9 - Présentation des différents rapports son papier	37
Figure 1.10 - Captures d'écran de l'affichage des métadonnées dans Pro Tools (Version 12.7.1).....	40

Chapitre II

Figure 2.1 - Détournement possible de l'enregistreur Nagra IV-S	51
Figure 2.2 - Présentation de la méthode de remise en phase	52
Figure 2.3 - Photographies des consoles analogiques de tournage.....	54
Figure 2.4 - Photographies des surfaces de contrôle de tournage	55
Figure 2.5 - Chaîne du son d'un signal microphonique avant et après la Conversion Analogique-Numérique (CAN) d'un enregistreur audionumérique	59
Figure 2.6 - Représentation des trois méthodes de modulation de volume	67

Chapitre III

Figure 3.1 - Présentation du filtrage Baxendall, semi-paramétrique et paramétrique	75
Figure 3.2 - Présentation des filtres à plateau	75
(à gauche Low Shelf ; à droite High Shelf).....	75
Figure 3.3 - Schéma représentant l'emplacement des données dans un enregistreur numérique	77
Figure 3.4 - Capture d'écran de la recherche de FsF dans Pro Tools	79
Figure 3.5 - Capture d'écran représentant le calcul séparant deux points d'automation dans Pro Tools	79
Figure 3.6 - Calcul du nombre de bits alloués pour la valeur temporelle relative	81
Figure 3.7 – Présentation des implémentations possibles pour la modulation de volume.....	83
Figure 3.8 - Tableau des paramètres de métadonnées de filtrage	84
Figure 3.9 - Tableau des paramètres des métadonnées liées à la phase.....	84
Figure 3.10 - Tableau des paramètres des métadonnées liées au panoramique	84
Figure 3.11 - Mise en forme en langage XML appliquée aux nouvelles métadonnées	86
Figure 3.12 - Évolution de la mise en forme des nouvelles métadonnées dynamiques.....	87
Figure 3.13 - Calculs théoriques du poids des nouvelles métadonnées dynamiques (modulation de volume).....	88
Figure 3.14 - Schéma présentant une possible évolution de l'échelle de gain	89
Figure 3.15 - Présentation du code Python pour créer un fichier test MIDI.....	94
Figure 3.16 - Présentation des branchements entre le potentiomètre et la carte Arduino	96
Figure 3.17 - Présentation de l'interface de contrôle dans Max 7	97
Figure 3.18 - Synoptique du fonctionnement des quatre étapes de la partie enregistrement du démonstrateur	99
Figure 3.19 - Présentation de l'enregistrement temporaire de la métadonnée Filtrage dans Max 7	101
Figure 3.20 - Capture d'écran du patcher n°4 dans Max 7.....	102
Figure 3.21 - Schéma fonctionnel du module de lecture du démonstrateur.....	104
Figure 3.22 - Courbes d'automation générées dans Pro Tools à partir du module de lecture.....	105

Table des annexes

Annexe A - Liste des tranches officielles disponibles dans les fichiers WAVE et BWF.....	116
Annexe B - Présentation de charte d'implémentation iXML de l'enregistreur SX-R4+ de Sonosax	118
Annexe C - Structure générale de la tranche iXML (version 2.02 datant du 20 mars 2018).....	119
Annexe D - Présentation du dictionnaire pour la balise <i>Function</i> de la tranche iXML.....	125
Annexe E - Tableau des métadonnées professionnelles de la tranche iXML.....	126
Annexe F - Questionnaires utilisés lors des entretiens avec les professionnels du son au cinéma	127
Annexe G - Chaîne du son dans un Cantar-X3 (présentée en page 70 du manuel)	130
Annexe H - Extrait du livre de Lucien Balibar	131
Annexe I - Tableau récapitulatif du filtrage employé dans les enregistreurs numériques et consoles de mixage analogiques actuels.....	132
Annexe J - Fichier XML n°1 généré à partir du démonstrateur	133
Annexe K - Fichier XML n°2 généré du démonstrateur.....	135
Annexe L - Caractéristiques du microcontrôleur <i>Arduino Uno</i>	136
Annexe M – Fichier XML utilisé pour générer les lignes d'automation dans Pro Tools.....	137

Annexe A - Liste des tranches officielles disponibles dans les fichiers WAVE et BWF

Voici une liste des autres tranches que l'on peut rencontrer dans un fichier WAVE :

- **<cue-ck>**

Cette tranche *Cue Points* contient des marqueurs ayant une position précise dans le fichier.

- **<playlist-ck>**

La tranche *PLAYLIST* permet de donner la position des marqueurs de la tranche *<cue-ck>* lors de la lecture du fichier.

- **<assoc-data-list>**

La tranche *Associated Data List* (en français Liste Associée aux Données) contient des noms sous forme de texte associés aux marqueurs de la tranche *<cue-ck>*. Cette tranche est directement liée au format IFF.

- **<smpl-ck>**

La tranche *Sample* permet d'accéder aux paramètres d'un instrument comme un *sampler* MIDI. On retrouve certaines correspondances avec la tranche *INSTRUMENT* du format AIFF.

- **<inst-ck>**

La tranche *Instrument* contient des informations du même ordre que la tranche *<smpl-ck>*

- **<qlty-ck>**

La tranche *Quality* est apparue en juillet 2001 lors du supplément n°2 de la spécification *Tech 3285*. Elle permet de renseigner l'ensemble des informations de la chaîne du son lors du transfert de données audio d'un support à un autre.

- **<levl-ck>**

La tranche *Peak Envelope* est apparue en juillet 2001 lors du supplément n°3 de la spécification *Tech 3285*. Elle informe sur les niveaux de crête du signal (par défaut sur des blocs de 256 échantillons) afin de faciliter l'échange des données audio entre les STAN.

- **<link-ck>**

La tranche *Link* dite « de liaison » est apparue en avril 2003 lors du supplément n°4 de la spécification *Tech 3285*. Les données de cette tranche sont stockées en langage XML. Certaines applications ne supportent pas les fichiers de plus de deux giga-octets, cette tranche permet de relier les différents éléments d'un même fichier.

- **<axml-ck>**

La tranche *aXML* est apparue en juillet 2003 lors du supplément n°5 de la spécification *Tech 3285*. Elle permet de renseigner en langage XML toutes les métadonnées développées dans les spécifications *Tech 3293*, « EBU Core Metadata Set for Radio Archives (CMSRA) », et *Tech 3295*, « EBU P_META Metadata Exchange Scheme ».

- **<dbmd-ck>**

La tranche *Dolby Metadata* est apparue en octobre 2009 lors du supplément n°6 de la spécification *Tech 3285*. Elle permet d'intégrer au format BWF les données liées aux différentes technologies développées par Dolby : Dolby E, Dolby Digital et Dolby Digital Plus.

- **<ISRC-ck>**

En 2012, l'AES a publié la spécification *Tech 3352*, « *The Carriage of Identifiers in the Broadcast Wave Format (BWF)* » pour implémenter l'ISRC, identifiant unique d'une œuvre enregistrée dans le domaine musicale, dans la tranche <axml-ck>.

- **ID3v2**

La tranche ID3v2 a été au départ créée pour les fichiers MPEG. Désormais, on peut écrire cette tranche dans un fichier WAVE mais, n'étant pas standardisé, on observe de nombreux problèmes d'incompatibilité entre les différents logiciels.

Annexe B - Présentation de charte d'implémentation iXML de l'enregistreur SX-R4+ de Sonosax



VENDOR MODEL		SONOSAX Recorders version 4.0 and greater		DATE	01.06.2011	
IXML MASTER TAG		IXML SUB TAGS		WRITTEN	READ	REMARKS
<PROJECT>				O	O	Max 8 chars
<SCENE>				O	O	Max 8 chars
<TAKE>				O	O	Range: 001..999
<TAPE>				O	O	Working day: YYYYMMDD
<CIRCLED>				O	O	
<WILD TRACK>				O	O	
<FALSE START>				O	O	
<NO GOOD>				O	O	
<FILE_UID>				O	O	= BWF Originator Reference
<UBITS>				X	X	
<NOTE>				O	O	Max 32 chars
<BEXT>				O	X	
<USER>				X	X	
<SPEED>				O	O	
<SPEED>	<NOTE>			X	X	
<SPEED>	<MASTER_SPEED>			X	X	
<SPEED>	<CURRENT_SPEED>			X	X	
<SPEED>	<TIMECODE_RATE>			O	O	
<SPEED>	<TIMECODE_FLAG>			O	O	
<SPEED>	<FILE_SAMPLE_RATE>			O	X	
<SPEED>	<AUDIO_BIT_DEPTH>			O	O	
<SPEED>	<DIGITIZER_SAMPLE_RATE>			O	O	
<SPEED>	<TIMESTAMP_SAMPLES_SINCE_MIDNIGHT_HI>			O	O	
<SPEED>	<TIMESTAMP_SAMPLES_SINCE_MIDNIGHT_LO>			O	O	
<SPEED>	<TIMESTAMP_SAMPLE_RATE>			O	X	
<SYNC_POINT_LIST>				X	X	
<SYNC_POINT>	<SYNC_POINT_TYPE>			X	X	
<SYNC_POINT>	<SYNC_POINT_FUNCTION>			X	X	
<SYNC_POINT>	<SYNC_POINT_COMMENT>			X	X	
<SYNC_POINT>	<SYNC_POINT_LOW>			X	X	
<SYNC_POINT>	<SYNC_POINT_HIGH>			X	X	
<SYNC_POINT>	<SYNC_POINT_EVENT_DURATION>			X	X	
<HISTORY>				X	X	
<HISTORY>	<ORIGINAL_FILENAME>			X	X	
<HISTORY>	<PARENT_FILENAME>			X	X	
<HISTORY>	<PARENT_UID>			X	X	
<FILE_SET>				O	O	
<FILE_SET>	<TOTAL_FILES>			O	O	
<FILE_SET>	<FAMILY_UID>			O	O	= BWF Originator Reference
<FILE_SET>	<FAMILY_NAME>			O	O	
<FILE_SET>	<FILE_SET_INDEX>			O	O	
<TRACK_LIST>				O	O	
<TRACK_LIST>	<TRACK_COUNT>			O	O	
<TRACK>	<CHANNEL_INDEX>			O	O	
<TRACK>	<INTERLEAVE_INDEX>			O	O	
<TRACK>	<NAME>			O	O	
<TRACK>	<FUNCTION>			X	X	
						O = YES
						X = NO
NOTES:						
Not all Sonosax recorders support track naming and NOTE						

Annexe C - Structure générale de la tranche iXML (version 2.02 datant du 20 mars 2018)

```

<?xml version="1.0" encoding="UTF-8"?>
<BWFXML>
  <IXML_VERSION>
  <PROJECT>
  <SCENE>
  <TAPE>
  <TAKE>
  <TAKE_TYPE>
  <NO_GOOD>
  <FALSE_START>
  <WILD_TRACK>
  <CIRCLED>
  <FILE_UID>
  <UBITS>
  <NOTE>
  <PRE_RECORD_SAMPLECOUNT>
  <SYNC_POINT_LIST>
    <SYNC_POINT_COUNT>
    <SYNC_POINT>
      <SYNC_POINT_TYPE>
      <SYNC_POINT_FUNCTION>
      <SYNC_POINT_COMMENT>
      <SYNC_POINT_LOW>
      <SYNC_POINT_HIGH>
      <SYNC_POINT_EVENT_DURATION>
    <SYNC_POINT>
      <SYNC_POINT_TYPE>
      <SYNC_POINT_FUNCTION>
      <SYNC_POINT_COMMENT>
      <SYNC_POINT_LOW>
      <SYNC_POINT_HIGH>
      <SYNC_POINT_EVENT_DURATION>
  <SPEED>
    <NOTE>
    <MASTER_SPEED>
    <CURRENT_SPEED>
    <TIMECODE_RATE>
    <TIMECODE_FLAG>
    <FILE_SAMPLE_RATE>
    <AUDIO_BIT_DEPTH>
    <DIGITIZER_SAMPLE_RATE>
    <TIMESTAMP_SAMPLES_SINCE_MIDNIGHT_HI>
    <TIMESTAMP_SAMPLES_SINCE_MIDNIGHT_LO>
    <TIMESTAMP_SAMPLE_RATE>
  <LOUDNESS>
    <LOUDNESS_VALUE>
    <LOUDNESS_RANGE>
    <MAX_TRUE_PEAK_LEVEL>
    <MAX_MOMENTARY_LOUDNESS>
    <MAX_SHORT_TERM_LOUDNESS>
  <HISTORY>
    <ORIGINAL_FILENAME>
    <PARENT_FILENAME>
    <PARENT_UID>

```

```

<FILE_SET>
  <TOTAL_FILES>
  <FAMILY_UID>
  <FAMILY_NAME>
  <FILE_SET_INDEX>
<TRACK_LIST>
  <TRACK_COUNT>
  <TRACK>
    <CHANNEL_INDEX>
    <INTERLEAVE_INDEX>
    <NAME>
    <FUNCTION>
  <TRACK>
    <CHANNEL_INDEX>
    <INTERLEAVE_INDEX>
    <NAME>
    <FUNCTION>
<BEXT>
  <BWF_DESCRIPTION>
  <BWF_ORIGINATOR>
  <BWF_ORIGINATOR_REFERENCE>
  <BWF_ORIGINATION_DATE>
  <BWF_ORIGINATION_TIME>
  <BWF_TIME_REFERENCE_LOW>
  <BWF_TIME_REFERENCE_HIGH>
  <BWF_VERSION>
  <BWF_RESERVED>
  <BWF_UMID>
  <BWF_CODING_HISTORY>
  <BWF_LOUDNESS_VALUE>
  <BWF_LOUDNESS_RANGE>
  <BWF_MAX_TRUE_PEAK_LEVEL>
  <BWF_MAX_MOMENTARY_LOUDNESS>
  <BWF_MAX_SHORT_TERM_LOUDNESS>
<USER>
  <FULL_TITLE>
  <DIRECTOR_NAME>
  <PRODUCTION_NAME>
  <PRODUCTION_ADDRESS>
  <PRODUCTION_EMAIL>
  <PRODUCTION_PHONE>
  <PRODUCTION_NOTE>
  <SOUND_MIXER_NAME>
  <SOUND_MIXER_ADDRESS>
  <SOUND_MIXER_EMAIL>
  <SOUND_MIXER_PHONE>
  <SOUND_MIXER_NOTE>
  <AUDIO_RECORDER_MODEL>
  <AUDIO_RECORDER_SERIAL_NUMBER>
  <AUDIO_RECORDER_FIRMWARE>
<LOCATION>
  <LOCATION_NAME>
  <LOCATION_GPS>
  <LOCATION_ALTITUDE>
  <LOCATION_TYPE>
  <LOCATION_TIME>

```

Suite



Groupe n°2 : les points de synchronisation

<SYNC_POINT_LIST>

<SYNC_POINT_COUNT>

<SYNC_POINT>

Le groupe *SYNC_POINT_LIST* est composé de deux balises principales : *SYNC_POINT_COUNT* et *SYNC_POINT*. La première indique le nombre (x) de points de synchronisation dans le fichier. La seconde est répétée (x) fois. Elle est constituée de six métadonnées.

<SYNC_POINT_TYPE>

<SYNC_POINT_LOW>

<SYNC_POINT_HIGH>

<SYNC_POINT_EVENT_DURATION>

La balise *SYNC_POINT_TYPE* utilise un dictionnaire avec les mots suivants : *RELATIVE* (valeur temporelle correspondant à une valeur depuis le début du fichier, par exemple, le nombre d'échantillons) et *ABSOLUTE* (valeur temporelle en *Timecode*).

Si le terme employé est *RELATIVE* alors la balise *SYNC_POINT_EVENT_DURATION* est référencée en nombre d'échantillons. Si le terme *ABSOLUTE* est employé alors les deux balises *SYNC_POINT_LOW* et *SYNC_POINT_HIGH* donnent un *Timecode* codé en échantillons.

<SYNC_POINT_FUNCTION>

Cette balise utilise un dictionnaire avec les mots suivants : *PRE_RECORD_SAMPLECOUNT* (*buffer* avant l'enregistrement), *SLATE_GENERIC* (peut servir pour l'auto-détection de *Slate*), *MARKER_GENERIC*, *MARKER_AUTOPLAY*, *MARKER_AUTOPLAYSTOP*, *MARKER_AUTOPLAYLOOP* et *GROUPE_OFFSET*.

L'auto-détection de clap par le Cantar-X permet d'accélérer la phase de conformation des *rushes* image et son avant le montage image.

<SYNC_POINT_COMMENT>

Il est aussi possible d'ajouter un commentaire pour décrire la fonction d'un *SYNC_POINT*.

En ce qui concerne la balise *SYNC_POINT*, l'AFSI a proposé d'ajouter deux sous-balises <SLATE> et <ANNOUNCEMENT>. La première serait décrite par un dictionnaire utilisant les mots suivants : HEAD_SLATE, TAIL_SLATE, NO_SLATE. La seconde utiliserait un autre dictionnaire : WITH, WITHOUT, ERROR.

Groupe n°3 : la fréquence image spécifique à une prise

<SPEED>

<NOTE>

<MASTER_SPEED>

Cette balise donne la valeur de la fréquence d'image du projet dans sa globalité. Elle n'est renseignée qu'une seule fois par le chef opérateur du son en début de tournage.

<CURRENT_SPEED>

Cette balise indique la valeur la plus importante de ce groupement de métadonnées car elle renseigne la fréquence d'échantillonnage utilisée pour la séquence ou le plan spécifique.

<TIMECODE_RATE>

La balise *TIMECODE_RATE* prend les valeurs suivantes de la fréquence d'image : 24, 25 et 29.97 images par seconde.

<TIMECODE_FLAG>

Elle indique s'il y a *Drop Frame*¹²⁰ (DF) ou non.

<FILE_SAMPLE_RATE>

<AUDIO_BIT_DEPTH>

Ces deux balises renseignent sur la fréquence d'échantillonnage et la quantification utilisée pour l'enregistrement du son.

¹²⁰ Le terme *Drop Frame* est un terme technique employé dans le domaine de l'image qui désigne l'action de palier, volontairement par « saut d'image », à un décalage temporel ayant inévitablement lieu à cause d'une différence de *Timecode* entre différentes machines.

<DIGITIZER_SAMPLE_RATE>

La balise *DIGITIZER_SAMPLE_RATE* indique la véritable fréquence d'échantillonnage utilisée au moment de la CAN du son. Par exemple, elle peut être égale à 48 048 Hz et non à 48 kHz.

<TIMESTAMP_SAMPLES_SINCE_MIDNIGHT_HI>

<TIMESTAMP_SAMPLES_SINCE_MIDNIGHT_LO>

Ces balises sont tous les deux codées sur 32 bits. Elles renseignent le *Timecode* en nombre d'échantillons après minuit.

<TIMESTAMP_SAMPLE_RATE>

Cette balise indique la fréquence d'échantillonnage utilisée pour fabriquer le *Timecode*.

Groupe n°5 : le nom des fichiers

<HISTORY>

<ORIGINAL_FILENAME>

<PARENT_FILENAME>

<PARENT_UID>

La balise *HISTORY* est composée de trois sous-balises et offre un lien de parenté entre les différents fichiers enregistrés simultanément. La sous-balise *ORIGINAL_FILENAME* correspond au nom original du fichier au moment de sa création tandis que la seconde *PARENT_FILENAME* permet de donner le nom du fichier polyphonique au moment de sa création. Enfin la dernière balise *PARENT_UID* permet de donner un paramètre supplémentaire correspondant à une forme de *TAPE ID* employé par la société Avid.

L'exemple le plus commun est la création d'un fichier polyphonique au moment de l'enregistrement et de l'emploi de fichiers monophoniques lors de la phase de post-production. Ces métadonnées permettent une meilleure gestion des *rushes* à l'étape de la conformation.

Groupe n°6 : les métadonnées de conformation

<FILE_SET>

La balise *FILE_SET* est composée de quatre sous-balises permettant de faciliter le lien entre des fichiers ayant été créés au même moment.

<TOTAL_FILES>

Cette sous-balise indique le nombre de fichiers créés à l'enregistrement.

<FAMILY_UID>

Cette balise reprend la valeur de la balise *FILE_UID* de la tranche iXML mais les quatre derniers caractères, correspondant au *Track Identifier*, sont mis à 0.

<FAMILY_NAME>

Cette balise reprend la valeur de la balise *ORIGINAL_FILENAME* en ne conservant pas le nom de la piste.

Les balises *FILE_UID* et *FAMILY_UID* sont employées pour la conformation automatique des *rushes* de tournage. Par exemple, Pyramix les utilise contrairement à Pro Tools qui emploie d'autres métadonnées (comme les informations de *Timecode*) et une donnée propriétaire le *Tape-ID* pour relier les fichiers de même lien de parenté.

<FILE_SET_INDEX>

Cette balise indique le nombre de pistes enregistrées en prenant en compte les pistes « non armées ».

Groupe n°7 : les métadonnées de piste

<TRACK_LIST>

<TRACK_COUNT>

<TRACK>

Le groupe *TRACK_LIST* est composé de deux balises principales : *TRACK_COUNT* et *TRACK*. La première sous-balise indique le nombre (x) de pistes dans le fichier. La seconde est répétée (x) fois. Elle est constituée de quatre métadonnées.

<CHANNEL_INDEX> <INTERLEAVE_INDEX>

La balise *CHANNEL_INDEX* permet de référencer une piste par rapport à sa piste originale tandis que la seconde *INTERLEAVE_INDEX* permet de numéroter la piste sans compter les pistes armées.

<NAME>
<FUNCTION>

La balise *NAME* donne le nom de la piste tandis que la balise *FUNCTION* permet de donner sa fonction en utilisant un dictionnaire présenté en *Annexe D*. Ces différentes fonctions permettent de spécifier la fonction d'un microphone spécifique dans une configuration de prise de son multicanal (LCR ou *Ambisonic*) ou stéréophonique (M-S, X-Y, etc.)

En fonction des enregistreurs, la gestion des noms de piste est différente. Par exemple, dans le Cantar-X3, il existe un répertoire nommé *Completion Track* regroupant jusqu'à cent références. Chaque nom appliqué à une piste est automatiquement enregistré dans ce répertoire afin de faciliter la réécriture de ce nom par la suite. Ce répertoire permet ainsi d'éviter les orthographes différentes pour un prénom ou nom d'un des personnages.

Groupe n°10 : la localisation

<LOCATION>

<LOCATION_NAME>
<LOCATION_GPS>
<LOCATION_ALTITUDE>

Ces trois premières balises indiquent respectivement le nom du lieu, les coordonnées GPS et l'altitude où les fichiers ont été enregistrés.

<LOCATION_TYPE>

Cette balise utilise un dictionnaire avec les mots suivants : INT (intérieur), EXT (extérieur) et INT/EXT. Elle permet de spécifier le type d'endroit où le son a été enregistré afin d'en faciliter le tri.

<LOCATION_TIME>

La balise *LOCATION_TIME* utilise un dictionnaire avec les mots suivants : SUNRISE, MORNING, MIDDAY, DAY, AFTERNOON, EVENING, SUNSET et NIGHT.

Annexe D - Présentation du dictionnaire pour la balise *Function* de la tranche iXML

Dans cette annexe, toutes les fonctions disponibles à travers la spécification du iXML ne sont pas présentées. Pour plus d'informations, toutes les informations sont disponibles sur le site officiel¹²¹ du iXML dans le menu *FUNCTION dictionary*.

Enregistrement avec couple stéréophonique

M-S

<FUNCTION>M-MID_SIDE</FUNCTION>
<FUNCTION>S-MID_SIDE</FUNCTION>

X-Y

<FUNCTION>X-X_Y</FUNCTION>
<FUNCTION>Y-X_Y</FUNCTION>

Quelconque

<FUNCTION>LEFT</FUNCTION>
<FUNCTION>RIGHT</FUNCTION>

Enregistrement avec prise de son multicanal

LCR

<FUNCTION>L-LCR</FUNCTION>
<FUNCTION>C-LCR</FUNCTION>
<FUNCTION>R-LCR</FUNCTION>

Quadriphonie

<FUNCTION>L-LCRS</FUNCTION>
<FUNCTION>C-LCRS</FUNCTION>
<FUNCTION>R-LCRS</FUNCTION>
<FUNCTION>S-LCRS</FUNCTION>

Ambisonic au format A*

<FUNCTION>W-SOUNDFIELD</FUNCTION>
<FUNCTION>X-SOUNDFIELD</FUNCTION>
<FUNCTION>Y-SOUNDFIELD</FUNCTION>
<FUNCTION>Z-SOUNDFIELD</FUNCTION>

*** Les formats B et de troisième ordre sont aussi disponibles dans la spécification.**

¹²¹ Disponible à l'adresse URL : <http://ixml.info/>

Annexe E - Tableau des métadonnées professionnelles de la tranche iXML

Groupe	Nom de Balise	Hiérarchie	Statique		Dynamique	
			Manuelle	Automatique	Manuelle	Automatique
-	<IXML_VERSION>	PROJET				
-	<PROJECT>	PROJET				
-	<SCENE>	PRISE				
-	<TAPE>	PRISE				
-	<TAKE>	PRISE				
-	<TAKE_TYPE>	PRISE				
-	<CIRCLED>	PRISE				
-	<FILE_UID>	PRISE				
-	<UBITS>	PRISE				
-	<NOTE>	PRISE				
<SYNC_POINT_LIST>	<SYNC_POINT_COUNT>	PRISE				
<SYNC_POINT_LIST> / <SYNC_POINT>	<SYNC_POINT_TYPE>	PRISE				
	<SYNC_POINT_FUNCTION>	PRISE				
	<SYNC_POINT_COMMENT>	PRISE				
	<SYNC_POINT_LOW>	PRISE				
	<SYNC_POINT_HIGH>	PRISE				
<SPEED>	<SYNC_POINT_EVENT_DURATION>	PRISE				
	<NOTE>	PRISE				
	<MASTER_SPEED>	PRISE				
	<CURRENT_SPEED>	PRISE				
	<TIMECODE_RATE>	PRISE				
	<TIMECODE_FLAG>	PRISE				
	<FILE_SAMPLE_RATE>	PRISE				
	<AUDIO_BIT_DEPTH>	PRISE				
	<DIGITIZER_SAMPLE_RATE>	PRISE				
	<TIMESTAMP_SAMPLES_SINCE_MIDNIGHT_HI>	PRISE				
	<TIMESTAMP_SAMPLES_SINCE_MIDNIGHT_LO>	PRISE				
	<TIMESTAMP_SAMPLE_RATE>	PRISE				
	<LOUDNESS_VALUE>	PRISE				
	<LOUDNESS_RANGE>	PRISE				
<LOUDNESS>	<MAX_TRUE_PEAK_LEVEL>	PRISE				
	<MAX_MOMENTARY_LOUDNESS>	PRISE				
	<MAX_SHORT_TERM_LOUDNESS>	PRISE				
	<ORIGINAL_FILENAME>	PRISE				
<HISTORY>	<PARENT_FILENAME>	PRISE				
	<PARENT_UID>	PRISE				

Groupe	Nom de Balise	Hiérarchie	Statique		Dynamique	
			Manuelle	Automatique	Manuelle	Automatique
<FILE_SET>	<TOTAL_FILES>	PRISE				
	<FAMILY_UID>	PRISE				
	<FAMILY_NAME>	PRISE				
	<FILE_SET_INDEX>	PRISE				
<TRACK_LIST>	<TRACK_COUNT>	PRISE				
	<CHANNEL_INDEX>	PISTE				
	<INTERLEAVE_INDEX>	PISTE				
	<NAME>	PISTE				
<TRACK_LIST> / <TRACK>	<FUNCTION>	PISTE				
	<BWF_DESCRIPTION>	PRISE				
	<BWF_ORIGINATOR>	PRISE				
	<BWF_ORIGINATOR_REFERENCE>	PRISE				
	<BWF_ORIGINATION_DATE>	PRISE				
	<BWF_ORIGINATION_TIME>	PRISE				
	<BWF_TIME_REFERENCE_LOW>	PRISE				
	<BWF_TIME_REFERENCE_HIGH>	PRISE				
	<BWF_VERSION>	PRISE				
	<BWF_RESERVED>	PRISE				
	<BWF_LUID>	PRISE				
	<BWF_CODING_HISTORY>	PRISE				
	<BWF_LOUDNESS_VALUE>	PRISE				
	<BWF_LOUDNESS_RANGE>	PRISE				
<BWF_MAX_TRUE_PEAK_LEVEL>	PRISE					
<BWF_MAX_MOMENTARY_LOUDNESS>	PRISE					
<BWF_MAX_SHORT_TERM_LOUDNESS>	PRISE					
<USER>	<FULL_TITLE>	PROJET				
	<DIRECTOR_NAME>	PROJET				
	<PRODUCTION_NAME>	PROJET				
	<PRODUCTION_ADDRESS>	PROJET				
	<PRODUCTION_EMAIL>	PROJET				
	<PRODUCTION_PHONE>	PROJET				
	<PRODUCTION_NOTE>	PROJET				
	<SOUND_MIXER_NAME>	PROJET				
	<SOUND_MIXER_ADDRESS>	PROJET				
	<SOUND_MIXER_EMAIL>	PROJET				
	<SOUND_MIXER_PHONE>	PROJET				
	<SOUND_MIXER_NOTE>	PROJET				
	<AUDIO_RECORDER_MODEL>	PROJET				
	<AUDIO_RECORDER_SERIAL_NUMBER>	PROJET				
<AUDIO_RECORDER_FIRMWARE>	PROJET					
<LOCATION>	<LOCATION_NAME>	PRISE				
	<LOCATION_GPS>	PRISE				
	<LOCATION_ALTITUDE>	PRISE				
	<LOCATION_TYPE>	PRISE				
	<LOCATION_TIME>	PRISE				

Annexe F - Questionnaires utilisés lors des entretiens avec les professionnels du son au cinéma

Questionnaire Chef opérateur du son

- Mixdown

- 1) À quoi correspond le mixdown lors de vos prises de son ? Que mettez-vous dans le *mixdown* Gauche et dans le *mixdown* Droit ?
- 2) Au moment de la prise de son, pour qui réalisez-vous le mixdown en premier lieu ? Quelle est sa fonction première ?
- 3) Ce mixdown est-il pour vous le représentant (garant) de la prise de son ? Doit-il être utilisé au maximum jusqu'au mixage du film ?

- Traitements

- 4) À quoi correspond le traitement fréquentiel (filtre et EQ) lors de la prise ?
- 5) À quoi correspond le traitement dynamique (mouvement de *fader*) lors de la prise ? Est-ce que vous laissez une modulation résiduelle ou vous fermez complètement le *fader* ?
- 6) Comment utilisez-vous le traitement de modulation de volume (pré ou *post fader* ou encore les *faders* contrôlant directement le gain de préamplification) ?
- 7) Cette modulation est-elle réalisée à chaque situation de tournage ou dépend-t-elle majoritairement du nombre de comédiens et de la préparation en amont de la séquence (improvisation) ?

- Métadonnées

- 8) Si ces informations (dynamiques et fréquentielles) étaient stockées sous forme de métadonnées, c'est-à-dire comme des lignes d'automatisation, seriez-vous prêt à les utiliser et à les fournir au monteur paroles ?
- 9) Changeront-elles votre manière de travailler au moment de la prise de son ?
- 10) Sachant qu'un grand nombre de personnes utilisent des consoles de mixage analogiques de type *Sonosax*, ces informations seraient alors irrécupérables. La proposition de ce mémoire pourrait-elle modifier l'emploi de ces mixettes ?
- 11) Modifieraient-elles la fonction de chef opérateur du son à l'avenir ?

Questionnaire Monteur paroles

- Mixdown

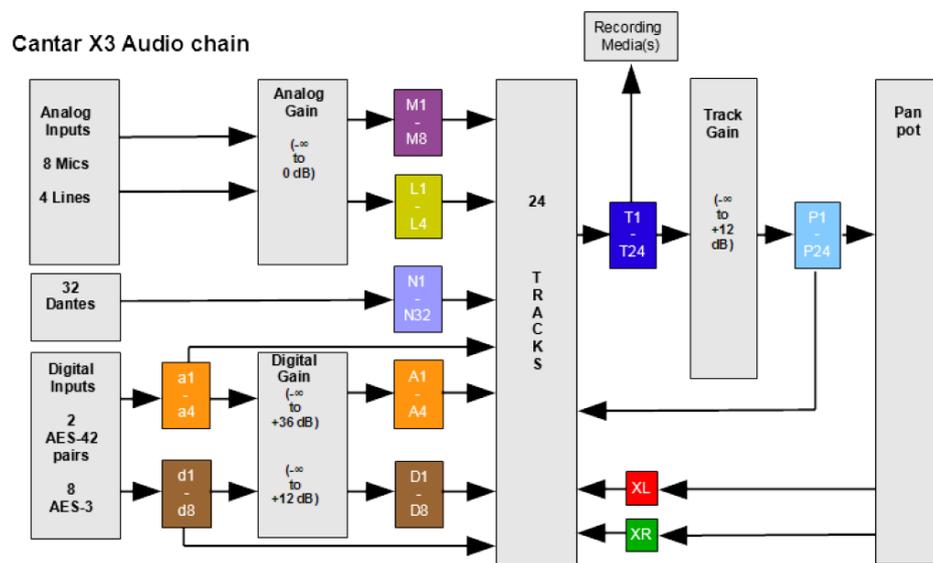
- 1) Dans la majorité des cas, le *mixdown* Gauche correspond à la perche en *post fader*, est-ce toujours le cas ? Si non, avez-vous des exemples précis de situations (deuxième perche, microphone en tant qu'appoint caché dans le décor, microphone(s) d'ambiance) ?
- 2) Dans la majorité des cas, le *mixdown* Droit correspond au mixage de tous les microphones cravates des comédiens présents sur la face, est-ce toujours le cas ? Si non, avez-vous des exemples précis de situations ?
- 3) Quelle attention portez-vous au mixdown délivré par le chef opérateur du son ? Est-ce que vous l'utilisez en premier lieu ? Pouvez-vous décrire précisément l'utilisation que vous en faites ?
- 4) Existe-t-il toujours ce « débat » entre l'utilisation du *mixdown* et des pistes dites divergées ?
- 5) En ce qui concerne les pistes divergées, comment abordez-vous le travail (tri, édition ou encore quinconce) ? Quel logiciel utilisez-vous pour trier les *rushes* de tournage ?
- 6) Quelle matière (*pré* ou *post fader*) préférez-vous recevoir lors du montage ?
- 7) Comment réalisez-vous le travail sur la dynamique des voix ? Utilisez-vous la fonction de *clip gain* dans Pro Tools afin de limiter dès le montage la dynamique des paroles ?

- Métadonnées

- 8) Au niveau du tri des *rushes* de tournage, quelles informations sous forme de métadonnées vous sont le plus utiles ?
- 9) Quel est l'utilité du rapport son version papier dans votre pratique ? Rencontrez-vous d'autres formats ?
- 10) Avez-vous des demandes particulières en ce qui concerne des informations manquantes provenant du tournage ?
- 11) Le travail du mémoire porte sur l'implémentation de nouvelles métadonnées : des métadonnées dites d'ingénierie. Elles correspondent à l'ensemble des traitements de dynamique (mouvement de *fader*) et fréquentiels (filtre et EQ) que réalise le chef opérateur du son au moment de la prise de son. Est-ce que ces informations semblaient-elles vous manquer lors de vos précédents montages parole ?
- 12) Pour le moment, la question se pose d'implémenter ces nouvelles métadonnées dans un fichier externe au fichier WAVE, cela vous semble-t-il pratique ?
- 13) À votre avis, implémentées sous forme de *Clip Effect*¹²², quelles fonctions permettraient-elles de combler, créer ou remplacer ?
- 14) Définies comme ci-dessus, ces nouvelles métadonnées correspondraient plutôt à l'étape de montage paroles ou à l'étape de mixage (pré-mixage parole) ?

¹²² Le terme *clip effects* correspond à un traitement audio non destructif dans Pro Tools depuis la version 12.6. Il se distingue d'un *plug-in* car il est appliqué spécifiquement à un *clip* et non à une piste entière.

Annexe G - Chaîne du son dans un Cantar-X3



Annexe H - Extrait du livre de Lucien Balibar

« Pourtant nous sommes tous confrontés à une sorte de dictature sournoise de la technologie, et certains intervenants de la post-production militent pour que la captation en direct reste la plus ouverte possible à une réinterprétation en montage-mixage ou en étalonnage. J'ai déjà par exemple entendu des demandes de mixeurs ou monteurs pour équiper systématiquement d'un micro HF tous les comédiens qui parlent, et de les laisser en permanence ouverts pendant les prises sur des pistes séparées, afin d'éviter un oubli dans les dialogues, ou de laisser le monteur paroles refaire entièrement l'équilibre sur son ordinateur. Je ne crois pas que cette méthode puisse devenir une norme satisfaisante. D'abord parce que la fabrication d'un film passe par des choix à effectuer au fur et à mesure des étapes de fabrication. Reporter intégralement les décisions vers l'étape suivante conduit d'une part à déresponsabiliser en amont et d'autre part à alourdir inutilement la charge de travail en aval. Ensuite parce que ce n'est absolument pas un gage de qualité de laisser les choses se dérouler sans intervenir dans l'instant et que malgré toutes ces pseudo-sécurités il y aura toujours des ratés, des imprévus, des mauvais choix dans le placement de tel ou tel micro. Le non-choix est une illusion de sécurité.

Peut-être que pour contenter tout le monde la technique nous apportera bientôt une solution qui semble intéressante : la possibilité sur les enregistreurs de tournage de proposer ce qui se trouve dans toutes les stations de montage virtuel ou dans les consoles de mixage automatisées, à savoir le fait d'enregistrer sur une couche logicielle séparée du son, des informations de volume que l'on peut retoucher et rejouer différemment. Ainsi par défaut le mélange livré par l'opérateur du son direct serait exploitable tel quel, mais il serait possible de changer ce mélange si besoin en intervenant sur les courbes de volume sans toucher au son original. Ceci dans le souci un peu idéaliste de contenter au mieux tous les techniciens de la chaîne sonore. »

BALIBAR, Lucien, *La chaîne du son au cinéma et à la télévision : De la prise de son à la post-production*, Paris, Dunod, coll. Audio-Photo-Vidéo, 2015, p. 157.

Annexe i - Tableau récapitulatif du filtrage employé dans les enregistreurs audionumériques et consoles de mixage analogiques actuels

Marque	Modèle	Version	EQ														
			HPF 1			LOW - 'LF (peaking ou shelving)						MID			HIGH		
			Fréquence [Hz]	Pente [dB/octave]	Fréquence [Hz]	Gain [dB]	Q	Note	Fréquence [Hz]	Gain [dB]	Q	Note	Fréquence [Hz]	Gain [dB]	Q	Note	
Aaton	Cantar-X3	2.23	60 à 180	-6 à -12	50 à 16 000	+/- 15	Fixe	Semi-paramétrique	50 à 16 000	+/- 15	0.2 à 1.0	Paramétric EQ	50 à 16 000	+/- 15	Fixe	High-Shelf	
	Cantar Mini		50	-18													
Aeta Audio Systems	4Mmix	1.7.6	50 à 200	-12													
			300	-6													
Cooper	CS 208 v2	2003	100	-6	100	+/- 12	Fixe	Passe-haut	500 à 5000	+/- 15	Fixe	Semi-paramétrique	10 000	+/- 12	Fixe	Passe-bas	
DCAudio	ITAK	2008	20 à 150	-12	100	+/- 9	Fixe	Graphique	300 à 5 000	+/- 10	Fixe	Semi-paramétrique	6500	+/- 9	Fixe	Graphique	
	Nagra VI	3.20	100	-12													
Nagra	Nagra Seven	2.10	30 à 300	-12 ou -24													
			30 à 300	-12 ou -24													
Sonosax	SX62R	4.1	25 à 350	-12													
	SX-R4+	1.2	60 ou 120	-12													
	SX-ST / SX-VT	2010	15 à 400	-18	80	+/-15	Fixe	Passe-haut	200 à 8 000	+/- 15	Fixe	Semi-paramétrique	8 000	+/- 15	Fixe	Passe-bas	
	SX-ES64	1.0	135	-6	80	+/-15	Fixe	Passe-haut	200 à 8 000	+/- 15	Fixe	Semi-paramétrique	8 000	+/- 15	Fixe	Passe-bas	
	SX-5	1999	20 à 500	-12	80	+/-15	Fixe	Passe-haut	200 à 8 000	+/- 15	Fixe	Semi-paramétrique	8 000	+/- 15	Fixe	Passe-bas	
Sound Devices	788T	3.02	40 à 320	-6 ou -12	30 à 20 000	+/-12	0.1 à 1.0										
	MixPre-3	1.0	40 à 160	-18													
	MixPre-6	1.11															
	MixPre-10T	1.51	80 à 240	-18	60 à 800	+/- 12	0.707 à 2	Paramétrique Low-Shelf	400 à 6 000	+/- 12	0.707 à 2	Paramétrique	8 000 à 16 000	+/- 12	0.707 à 2	Paramétrique Low-Shelf	
	688	4.51	50	-6													
Zaxcom	664	4.51	80 à 240	-6 ou -12													
	552																
	Deva 5.8	7.52	30 à 240	-	30 à 20 000	+/- 24	0.5 à 9.9	Paramétrique	30 à 20 000	+/- 24	0.5 à 9.9	Paramétrique	30 à 20 000	+/- 24	0.5 à 9.9	Paramétrique	
Zoom	Deva 24	1.42	0 à 240	-	30 à 20 000	+/- 24	Fixe	Low-Shelf	30 à 20 000	+/- 24	0.5 à 9.9	Paramétrique	30 à 20 000	+/- 24	Fixe	High-Shelf	
	Nomad MAXX	7.77 2.13	40 à 230	-													
Zoom	F8	-	10 et 240	-													


```
<Piste_4>
  <Filtre1>
    <type> highpass </type>
    <Fcoupure> 50. </Fcoupure>
    <Pente> - 6 dB/octave </Pente>
  </Filtre1>
</Piste_4>
</Filtrage>

<Phase>
  <Piste_3> NON , 0 ; </Piste_3>
  <Piste_4> OUI , 0 ; </Piste_4>
</Phase>

<Panoramique>
  <Piste_3> 0 ; 100 </Piste_3>
  <Piste_4> 0 ; 100 </Piste_4>
</Panoramique>

</Traitements>
```

Annexe K - Fichier XML n°2 généré du démonstrateur

<?xml version="1.0" encoding="UTF-8"?>

<Traitements>

<Modulation>

<Piste_3>

-28.4, 0 ; -28.4, 0 ; -13.7, 14784 ; -15.6, 14784 ; -17.4, 14784 ; -19.3, 20416 ; -20.5, 20416 ; -21.8, 20416 ; -23.0, 20416 ; -24.9, 24512 ; -27.4, 24512 ; -29.2, 24512 ; -31.7, 24512 ; -34.8, 29632 ; -37.9, 29632 ; -41.0, 29632 ; -44.7, 29632 ; -47.2, 33728 ; -50.3, 33728 ; -53.4, 33728 ; -54.7, 33728 ; -56.5, 39360 ; -57.8, 39360 ; -59.6, 39360 ; -62.7, 39360 ; -65.2, 43456 ; -67.7, 43456 ; -71.4, 43456 ; -73.9, 43456 ; -76.4, 49088 ; -78.9, 49088 ; -80.7, 49088 ; -82.0, 49088 ; -84.4, 49088 ; -86.3, 53184 ; -88.8, 53184 ; -89.4, 53184 ; -91.3, 53184 ; -92.5, 58304 ; -94.4, 58304 ; -97.5, 58304 ; -99.9, 58304 ; -102.4, 62400 ; -103.0, 62400 ; -104.3, 62400 ; -104.9, 62400 ; -104.9, 68032 ; -102.4, 68032 ; -100.2, 201664 ; -98.3, 207296 ; -95.9, 207296 ; -90.3, 207296 ; -85.9, 207296 ; -82.8, 212928 ; -78.5, 212928 ; -76.6, 212928 ; -75.4, 212928 ; -74.1, 217024 ; -71.0, 217024 ; -69.2, 217024 ; -66.7, 217024 ; -64.2, 222144 ; -61.1, 222144 ; -56.8, 222144 ; -52.4, 222144 ; -49.9, 226240 ; -48.1, 226240 ; -46.2, 226240 ; -45.0, 226240 ; -43.1, 231872 ; -41.3, 231872 ; -38.8, 231872 ; -37.5, 235968 ; -36.3, 235968 ; -36.3, 241600 ; -37.5, 241600 ; -38.8, 241600 ; -41.9, 245696 ; -45.6, 245696 ; -49.9, 245696 ; -53.1, 245696 ; -55.5, 250816 ; -57.4, 250816 ; -58.6, 250816 ; -60.5, 250816 ; -61.7, 254912 ; -63.0, 254912 ; -64.2, 254912 ; -63.6, 260544 ; -53.7, 260544 ; -45.6, 264640 ; -41.9, 264640 ; -41.9, 278464 ; -45.0, 278464 ; -48.1, 278464 ; -51.2, 278464 ; -55.5, 278464 ; -59.9, 283584 ; -63.0, 283584 ; -66.7, 283584 ; -69.8, 283584 ; -71.7, 289216 ; -73.5, 289216 ; -75.4, 289216 ; -77.2, 289216 ; -79.1, 293312 ; -81.0, 293312 ; -82.8, 293312 ; -85.9, 298944 ; -87.8, 298944 ; -90.3, 298944 ; -92.8, 298944 ; -95.2, 303040 ; -96.5, 303040 ; -97.7, 303040 ; -98.3, 303040 ; -98.3, 308160 ; -98.3, 308160 ; -97.1, 308160 ; -59.5, 394176 ; -58.9, 394176 ; -55.8, 399808 ; -55.2, 399808 ; -53.9, 403904 ; -51.4, 409536 ; -49.6, 409536 ; -47.1, 409536 ; -44.6, 409536 ; -42.7, 418752 ; -41.5, 418752 ; -40.9, 418752 ; -39.0, 422848 ; -37.8, 422848 ; -36.5, 422848 ; -34.7, 422848 ; -33.4, 428480 ; -32.8, 432576 ; -32.2, 432576 ; -33.4, 442304 ; -34.7, 442304 ; -36.5, 442304 ; -39.0, 442304 ; -41.5, 442304 ; -42.7, 447424 ; -44.0, 447424 ; -44.6, 451520 ; -45.8, 451520 ; -47.1, 457152 ; -48.3, 457152 ; -49.6, 457152 ; -50.8, 461248 ; -52.1, 461248 ; -52.7, 461248 ; -54.5, 466880 ; -56.4, 466880 ; -57.0, 466880 ; -57.6, 470976 ; -58.9, 470976 ; -59.5, 470976 ; -60.7, 470976 ; -61.4, 470976 ; -62.0, 476096 ; -63.2, 476096 ; -64.5, 476096 ; -65.1, 480192 ; -66.3, 480192 ; -67.6, 480192 ; -68.8, 485824 ; -69.4, 485824 ; -70.0, 485824 ; -71.3, 489920 ; -72.5, 495552 ; -72.5, 495552 ; -69.4, 504768 ; -63.2, 504768 ; -58.3, 508864 ; -55.2, 508864 ; -52.1, 508864 ; -50.2, 508864 ; -49.0, 514496 ; -47.7, 514496 ; -46.5, 514496 ; -43.4, 514496 ; -41.5, 520128 ; -39.6, 520128 ; -37.8, 520128 ; -35.9, 520128 ; -34.1, 524224 ; -32.2, 524224 ; -31.6, 524224 ; -29.7, 524224 ; -27.9, 528320 ; -27.2, 528320 ; -26.6, 533440 ; -26.6, 533440 ; -27.2, 533440 ; -31.0, 533440 ; -35.3, 533440 ;

</Piste_3>

<Piste_4>

-51.0, 0 ; -51.0, 0 ; -53.1, 111552 ; -52.5, 111552 ; -48.2, 111552 ; -42.6, 115648 ; -39.5, 115648 ; -37.0, 115648 ; -36.4, 115648 ; -35.8, 121280 ; -34.5, 121280 ; -33.3, 121280 ; -31.4, 121280 ; -30.2, 125376 ; -28.9, 125376 ; -27.7, 125376 ; -26.5, 131008 ; -25.8, 131008 ; -24.6, 131008 ; -25.8, 139200 ; -28.9, 139200 ; -32.0, 139200 ; -36.4, 144320 ; -39.5, 144320 ; -42.6, 144320 ; -45.1, 144320 ; -48.8, 149952 ; -53.1, 149952 ; -57.5, 149952 ; -62.4, 149952 ; -66.2, 154048 ; -69.3, 154048 ; -71.1, 154048 ; -72.4, 154048 ; -73.6, 159680 ; -76.1, 159680 ; -78.6, 159680 ; -81.1, 159680 ; -83.5, 163776 ; -86.0, 163776 ; -89.7, 163776 ; -93.5, 163776 ; -95.9, 168896 ; -99.0, 168896 ; -102.8, 168896 ; -106.5, 168896 ; -109.6, 172992 ; -114.6, 172992 ; -117.7, 172992 ; -119.5, 172992 ; -120.8, 178624 ; -122.0, 178624 ; -123.9, 178624 ; -125.7, 178624 ; -126.3, 184256 ; -126.3, 184256 ; -53.8, 336832 ; -52.5, 336832 ; -47.6, 340928 ; -44.5, 340928 ; -42.0, 340928 ; -40.1, 340928 ; -38.9, 346560 ; -36.4, 346560 ; -35.2, 346560 ; -33.3, 346560 ; -31.4, 352192 ; -29.6, 352192 ; -28.3, 352192 ; -29.6, 356288 ; -33.3, 360384 ; -36.4, 360384 ; -39.5, 360384 ; -42.0, 360384 ; -43.8, 365504 ; -45.1, 365504 ; -46.3, 365504 ; -47.6, 365504 ; -48.8, 365504 ; -51.9, 369600 ; -53.2, 369600 ; -54.4, 369600 ; -55.0, 375232 ; -55.6, 375232 ; -56.3, 375232 ; -58.7, 375232 ; -60.6, 375232 ; -61.8, 380864 ; -63.7, 380864 ; -66.2, 380864 ; -66.8, 380864 ; -68.8, 380864 ; -71.7, 562112 ; -73.5, 562112 ; -75.2, 562112 ; -76.9, 567744 ; -78.5, 567744 ; -80.4, 567744 ; -82.5, 567744 ; -84.7, 567744 ; -87.1, 571840 ; -89.5, 571840 ; -91.9, 571840 ; -94.3, 577472 ; -96.7, 577472 ; -99.1, 581568 ; -101.5, 581568 ; -103.9, 581568 ; -106.3, 585664 ; -108.7, 585664 ; -111.1, 585664 ; -113.5, 590784 ; -115.9, 590784 ; -118.3, 590784 ;

</Piste_4>

</Modulation>

<Phase>

<Piste_3> NON , 0 ; </Piste_3>

<Piste_4> NON , 0 ; </Piste_4>

</Phase>

<Panoramique>

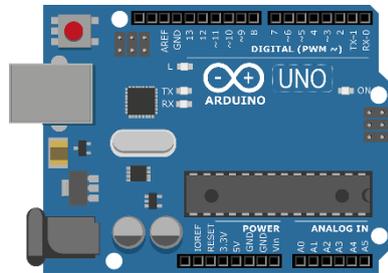
<Piste_3> 0 ; 100 </Piste_3>

<Piste_4> 0 ; 100 </Piste_4>

</Panoramique>

</Traitements>

Annexe L - Caractéristiques du microcontrôleur *Arduino Uno*



Microcontroller	ATmega328P
Operating Voltage	5V
Input Voltage (recommended)	7-12V
Input Voltage (limit)	6-20V
Digital I/O Pins	14 (of which 6 provide PWM output)
PWM Digital I/O Pins	6
Analog Input Pins	6
DC Current per I/O Pin	20 mA
DC Current for 3.3V Pin	50 mA
Flash Memory	32 KB (ATmega328P) of which 0.5 KB used by bootloader
SRAM	2 KB (ATmega328P)
EEPROM	1 KB (ATmega328P)
Clock Speed	16 MHz
LED_BUILTIN	13
Length	68.6 mm
Width	53.4 mm
Weight	25 g

École nationale supérieure Louis-Lumière 2018