Le Codage MPEG Surround Et Ses Applications À Une Diffusion Télévisuelle

Mémoire de fin d'études

Annabelle NEGRO

Section Son – Promotion 2013

Rédigé sous la direction de M. Claude GAZEAU et M. Manuel NAUDIN

Jury de soutenance encadré par M. Alan Blum

Mémoire soutenu le 19 Juin 2013

Louis Lumière
école nationale supérieure

REMERCIEMENTS

Je voudrais tout d'abord remercier mes deux directeurs de mémoire : M. Claude GAZEAU et M. Manuel NAUDIN.

Merci à Claude pour ses précieux conseils, sa confiance et son aide durant ces trois années à l'École. Merci à lui, ainsi qu'à la société Neyrac Films d'avoir accepté de me prêter des extraits du téléfilm *Jusqu'à l'enfer* pour mes tests.

Merci à Manuel pour son implication tout au long de ce mémoire, son accueil dans les locaux du service Innovations & Développement à France Télévisions, son aide, sa disponibilité, ses conseils avisés, le temps qu'il m'a consacré, sa patience, la confiance qu'il m'a accordée, ainsi que pour ses relectures. Je voudrais également le remercier de m'avoir fourni le documentaire *La vie moderne* de Raymond Depardon, ainsi que l'épisode de la série *U.S. Marshals*.

Merci à M. Alan BLUM, mon rapporteur, pour ses cours de Techniques Audio qui m'ont beaucoup aidée dans la rédaction de ce mémoire, pour l'intérêt qu'il a porté à mon sujet et pour sa participation aux tests perceptifs.

Je voudrais remercier chaleureusement toutes les équipes de France Télévisions, et plus particulièrement :

♪ L'équipe du service Innovations & Développement pour son accueil et la mise à disposition du laboratoire et de la licence du plugin Fraunhofer Pro-Codec de Sonnox : merci à Matthieu PARMENTIER, Lidwine HÔ, Claire MERIENNE-SANTONI, Olivier JOUINOT et Sandrine MARTY,

♪ M. Éric JOURNOUX, qui m'avait permis de faire un stage dans les régies de France 2 et France 3 et m'a donné envie de faire un mémoire en lien avec la télévision, ainsi que pour son aide lorsque je l'ai sollicité,

♪ M. Philippe DUMAS et M. Gilles PLAIDEUX pour leurs précieux renseignements concernant la diffusion des chaînes de télévision.

Merci également à M. Jean CHATAURET, ainsi qu'à M. David KULAS de la production François Roussillon & Associés, pour m'avoir autorisée à utiliser quelques extraits de l'opéra filmé *L'Italiana in Algeri* pour mes tests perceptifs.

Un grand merci à tous les participants à mes tests perceptifs, qu'ils soient étudiants à l'école, professionnels du son ou musiciens, ils m'ont accordé leur temps, et m'ont rendu un grand service en acceptant de se soumettre à ces tests. : Manuel N., Matthieu P., Tristan G., Pierre B., Antoine G., François H., Léa C., Adrien L., Léo G., Frédéric D., Alan B., Claire M., Philippe V., Alix M., Olivier J., Jonathan S., Matthieu T., Philippe K., Camille C., Christian P., Éric M., Baptiste P., Sandrine M., Éric J., François M., Larbi A., Yoann V., Lidwine H., Éric M., Rémi C., et Jean-François H..

Je remercie aussi Messieurs Andreas Hölzer et Alfonso Carrera de la société Fraunhofer IIS, pour leurs réponses à mes questions concernant le codage MPEG Surround et le plugin de Sonnox.

Je tiens également à remercier les enseignants de l'École Nationale Supérieure Louis Lumière pour leur disponibilité et leurs enseignements tout au long de ses trois ans, ainsi que M. Etienne Hendrickx pour ses conseils en analyse statistique, Mme Agnès Hominal pour son aide logistique et M. Florent Fajole pour ses recommandations bibliographiques.

Je souhaite également remercier Peter NAYLOR pour nos longues journées de mixage durant plusieurs mois, nos échanges, et pour sa contribution à la rédaction du résumé en anglais.

Enfin, je voudrais remercier mes proches : merci à David, ainsi qu'à mes parents, à ma sœur et à ma grand-mère pour leur soutien sans faille tout au long de ces trois dernières années. Merci Maman pour les nombreuses heures de relecture.

RÉSUMÉ

Depuis sa naissance dans les années 1930 avec une seule chaîne en noir et blanc et seulement quelques heures de programmes par jour, la télévision a beaucoup évolué et notamment durant cette dernière décennie avec l'arrivée de la télévision numérique terrestre, l'arrêt de la diffusion analogique, le développement de la télévision par les fournisseurs d'accès à Internet, et tous les services désormais proposés : télévision de rattrapage, vidéo à la demande, télévision connectée, etc.

On peut désormais regarder la télévision sur différents supports : écran plat, ordinateur, tablette numérique, smartphone. La télévision numérique terrestre haute définition supporte la diffusion d'une image haute définition et de l'audio multicanal (5.1), tandis que la diffusion en définition standard ne permet qu'une image en définition standard et un son stéréophonique. De plus, les chaînes fournissent aussi leur signal à des fournisseurs d'accès à Internet, ainsi qu'à des sous-traitants qui gèrent les services de télévision de rattrapage. La multiplication des supports de diffusion oblige alors la cohabitation de différents formats, tant pour l'image que pour l'audio.

Le MPEG Surround est un format de codage audio multicanal, de type « surround-parametric », qui permet d'encoder un signal multicanal 5.1 en un flux audio encapsulé en MPEG-4, qui contient un signal audio stéréophonique, ainsi qu'un flux de données qui renferme les informations de spatialisation. Ce format permet alors une diffusion audio multicanale sur des vecteurs limités en bande passante, tels que la télévision numérique terrestre en définition standard, le streaming et la vidéo à la demande. Ce format garantit une parfaite compatibilité 5.1/stéréo sur la plupart des

lecteurs multimédias grand public, qui pourrait éviter une double diffusion, ainsi qu'une compatibilité binaurale.

Je vais débuter ce mémoire par un rapide état des lieux du son à la télévision, à savoir les formats de diffusion actuels, les différents services proposés, et les équipements audio des Français. Ensuite, je rappellerai les mécanismes de perception auditive humaine des sons spatialisés. Puis, je présenterai plusieurs aspects du codage MPEG Surround : les étapes d'encodage et de décodage, les algorithmes utilisés, les paramètres réglables, ainsi que le fonctionnement du plugin Fraunhofer Pro-Codec de Sonnox. Enfin, je détaillerai mon protocole de tests perceptifs, qui devra permettre de déterminer les paramètres de codage satisfaisants du MPEG Surround (débit, format), ainsi que l'impact des étapes d'encodage et de décodage sur les paramètres de la recommandation R128, que tous les programmes télévisés doivent dorénavant respecter. Je devrai donc établir si sa qualité est suffisante et s'il pourrait être utilisé dans la diffusion audio de programmes télévisés dans les années à venir.

Mots clés : MPEG Surround. Multicanal. Diffusion. Encodage. Télévision.

ABSTRACT

Since its birth in the 30's with just one black and white channel and only a few hours of programmes every day, television has progressed rapidly, in particular during the last decade with the arrival of digital terrestrial television, analogue broadcasting switch-off, the development of television services by the Internet Access Providers, plus all the other services now offered: replay television, video on demand, connected television, etc.

Television can now be watched on different platforms: flat-screen TV, computer, touchpad and smartphone. High definition digital terrestrial television allows the broadcast of a high definition image and multi-channel audio (5.1), whereas standard definition broadcast allows only for a standard definition image and stereophonic sound. Moreover, the TV channels also provide their signal feed to the Internet Access Providers, and to subcontractors who handle replay TV services. This multiplication of methods of diffusion requires cohabitation between the different formats, for image as much as sound.

MPEG Surround is a multi-channel audio coding format, a type of « surround parametric » that allows a multi-channel 5.1 signal to be encoded in an MPEG-4 encapsulated bit-stream, which contains an audio stereophonic signal, as well as a data stream containing the spatialization data. This format allows for a multi-channel audio broadcast on low-bandwidth vectors, such as standard definition digital television, streaming, and video on demand. This format guarantees a perfect 5.1 multi-channel/stereo compatibility on most consumer media players, which is able to avoid a double broadcast, and maintain binaural compatibility.

I will begin this Master's thesis by a short review of sound in television, bearing in mind the current broadcasting formats, the different services offered and French audio equipment. Then, I will refer to the mechanisms of human audio perception for spatialized sound. Next, I will present several features of the MPEG Surround coding: the coding and decoding stages, the algorithms used, the adjustable parameters, as well as the functioning of the Sonnox Fraunhofer Pro-Codec plugin. Finally, I will explain my protocol for the perceptive tests in detail, which must be able to determine satisfactory MPEG Surround parameters (bitrate, formats), as well as the impact of the coding and decoding stages on the parameters of the R128 recommendation, which all televised programmes must observe from now on. I will have to determine if the audio quality is sufficient and if MPEG Surround could be used for audio in televised programmes broadcast over the next few years.

Key words: MPEG Surround. Multi-channel. Broadcast. Coding. Television.

SOMMAIRE

« Fais de ta vie un rêve, et d'un rêve, une réalité. »

Antoine de Saint-Exupéry

INTRODUCTION

La télévision est née au début du XXème siècle, mais il a fallu attendre 1935 pour voir l'apparition de la première chaîne de télévision française en noir et blanc, qui émettait seulement quelques heures de programmes par jour. La télévision a connu de grandes mutations au fil des décennies, notamment en 1967 avec l'apparition de la couleur. Tout d'abord perçue comme une grande avancée technologique réservée aux ménages les plus aisés, la télévision s'est progressivement installée dans tous les foyers au cours du temps, bouleversant le quotidien des Français. Au fil des ans, la télévision a beaucoup évolué, avec la multiplication des chaînes gratuites, l'apparition de chaînes payantes, les évolutions des technologies de transmission, la diffusion de programmes vingt-quatre heures sur vingt-quatre, sept jours sur sept. Au cours des années 2000, la télévision a connu de profondes métamorphoses, avec notamment l'arrivée de la télévision numérique terrestre (TNT), l'arrêt de la diffusion analogique hertzienne en 2011, la télévision sur IP[1], et l'apparition de nouveaux modes de consommation : la télévision connectée, qui permet d'accéder à des contenus bonus avec Internet, la télévision de rattrapage, la vidéo à la demande, etc. Le son à la télévision a connu lui aussi de grandes innovations, tout d'abord monophonique aux débuts de la télévision, puis stéréophonique, et désormais multicanal et multi-linguiste.

La télévision est avant tout un média convivial, qui peut être tour à tour informatif, divertissant, ou encore éducatif. La multiplication des chaînes permet de proposer un large choix de programmes (information, fictions, magazines, talk-shows, jeux télévisés, documentaires, spectacle vivant, événements sportifs, etc.), diffusés par des

[1] La télévision sur IP ou IPTV est un vecteur de diffusion de la télévision via un réseau utilisant le protocole IP (Internet Protocol). Ce type de diffusion sera détaillée dans le chapitre 1, 2.1.4 page 30.

chaînes généralistes (par exemple TF1, France 2, France 3, M6, etc.), ou par des chaînes thématiques (chaînes sportives, chaînes d'information, chaînes cinéma, chaînes musicales, chaînes dédiées aux enfants, etc.) : ainsi, tout un chacun, quelque soit l'âge ou les centres d'intérêt, peut trouver des programmes qui l'intéresse.

Évidemment, on peut reprocher beaucoup de défauts à la télévision, tant sur le plan technique que sur le plan artistique : les concepts de programmes souvent semblables d'une chaîne à l'autre, le manque d'originalité de certains programmes, les innombrables publicités, la diffusion de bons programmes à des horaires trop tardifs, les désynchronisations image et son, les nombreux play-backs dans les programmes musicaux, ou encore les rediffusions très fréquentes. Néanmoins, il faut reconnaître que la télévision apporte au sein des foyers et de façon quasi-gratuite (si l'on excepte la redevance télévisuelle payée par tous, et les chaînes payantes) du divertissement, de la culture, de l'information et des évènements sportifs. La télévision n'a pas vocation à remplacer ni le spectacle vivant, ni le cinéma, ni un événement sportif en « live », mais permet à des personnes qui n'en auraient pas toujours les moyens de suivre et de profiter de ces évènements à travers leur petit écran[2]. De plus, intégrée au sein des foyers, la télévision rassemble des personnes devant un même programme, qui peuvent alors échanger leurs impressions, critiquer, communiquer à propos de sujets liés à ce qu'ils sont en train de regarder. On trouve aussi des programmes de qualité, dont les concepts sont innovants et qui nécessitent qu'on soigne la diffusion pour transmettre la meilleure qualité possible. De plus, les modes de consommation ayant beaucoup évolué, les chaînes proposent désormais des services complémentaires (télévision de rattrapage, vidéo à la demande, télévision connectée avec des programmes interactifs), et la télévision peut dorénavant être regardée sur d'autres supports (par exemple sur l'ordinateur, sur le smartphone ou sur une tablette), et dans différents lieux (dans la rue, dans les transports en commun). La télévision n'est donc plus exclusivement réservée au salon des foyers, et l'écoute sur ces différents supports est par conséquent totalement disparate.

[2] Petit écran : expression familière désignant la télévision, par opposition au « grand écran » qui désigne le cinéma.

Aujourd'hui, la télévision peut s'écouter avec les haut-parleurs du téléviseur, avec deux enceintes externes, avec un système home-cinéma (plus ou moins bien installé d'ailleurs), ou encore au casque. Tandis que les chaînes effectuent une double diffusion audio pour la télévision numérique en définition standard (son stéréophonique) et en haute définition (son multicanal), que l'audio est souvent ré-encodé et compressé par les fournisseurs d'accès à Internet ou par les sous-traitants pour la télévision de rattrapage, et considérant les conditions très hétérogènes d'écoute, il n'est donc pas évident de trouver un format et un débit qui conviennent à tous les usages.

A la suite de mon premier stage dans les régies de France 2 et France 3 en 2011, j'ai eu envie d'en apprendre davantage sur les secrets de fabrication et de diffusion des chaînes de télévision. Dès lors, je réfléchissais à un sujet de mémoire technique qui se rattachait à ce domaine.

En Février 2012, j'ai rencontré M. Manuel Naudin, qui allait devenir mon directeur externe de mémoire. Après de riches échanges, j'ai choisi d'étudier le codage MPEG Surround et ses possibles applications à une diffusion télévisuelle.

Le codage MPEG Surround a été développé par la société Fraunhofer IIS, puis normalisé en 2007 par le Moving Picture Expert Group. Dans sa plus simple implémentation, ce codage réduit un signal multicanal 5.1 en un signal stéréophonique, accompagné d'un flux de données de spatialisation, extraites du signal 5.1. Le downmix stéréophonique est encodé dans un codeur principal, dans le format Advanced Audio Coding Low Complexity (AAC-LC) ou dans le format High-Efficiency Advanced Audio Coding (HE-AAC) ou plus rarement en MPEG 1 Layer 2. Ce downmix, encodé, et le flux de données de spatialisation sont alors encapsulés dans un flux MPEG-4, et c'est ce flux qui serait diffusé. Les principaux intérêts de ce codage sont la réduction de débit et la

compatibilité avec la majorité des décodeurs, ce qui évite une double diffusion. En effet, lors de la réception du flux MPEG-4, deux options se présentent. Soit le décodeur est compatible MPEG Surround, le signal 5.1 est reconstruit à partir du downmix stéréophonique et du flux de données de spatialisation. Sinon, si le récepteur n'est pas compatible, le signal est décodé comme une stéréo. Ce codage est d'ailleurs évolutif, puisqu'il supporte des signaux multicanaux qui peuvent comporter jusqu'à vingt-sept canaux : on peut alors imaginer transporter un signal audio 22.2 encodé en un signal 5.1, associé à un flux de données de spatialisation.

Ce codage apparaît prometteur, puisque très efficace en matière de débit, il pourrait alors permettre une diffusion audio multicanale sur des vecteurs limités en bande-passante, et où on ne peut envisager une diffusion en canaux discrets, comme par exemple en télévision numérique terrestre en définition standard, ou encore pour les services de streaming ou de vidéo à la demande. De plus, ce codage est compatible en binaural : à partir du flux de données de spatialisation, le décodeur MPEG Surround est capable de reconstruire un signal binaural simulant une écoute multicanale. Cette compatibilité permettrait d'écouter un programme télévisé en binaural au casque sur une tablette ou un smartphone, à condition de posséder une petite application possédant un décodeur MPEG Surround qui intègre un encodeur binaural, auquel on puisse ajouter nos propres HRTF.

Je vais donc débuter ce mémoire par un rapide historique de la télévision, et détailler ce qu'est la télévision aujourd'hui : les vecteurs de diffusion, les nouveaux modes de consommation de ce média, les supports d'écoute et de visionnage. Le deuxième chapitre sera consacré aux phénomènes psycho-acoustiques de l'oreille humaine (effets de masque, perception de l'espace sonore, etc.). Ensuite, je détaillerai les principaux formats actuels de diffusion audio à la télévision. Dans le quatrième chapitre, je décrirai les étapes d'encodage et de décodage du MPEG Surround, ainsi que les résultats de tests perceptifs effectués par d'autres organismes. Le cinquième chapitre précisera les protocoles de mes tests perceptifs, ainsi que le fonctionnement du plugin Fraunhofer Pro-Codec de Sonnox, utilisé pour réaliser mes encodages et mes décodages, et contiendra l'analyse statistique

des résultats. Enfin, le dernier chapitre sera dédié à un rappel des paramètres de la recommandation R128, et aux résultats des mesures de loudness effectuées sur trois programmes télévisés intégraux, encodés puis décodés, afin d'observer l'impact du codage MPEG Surround sur les valeurs loudness. La conclusion de ce mémoire devra déterminer les avantages et les inconvénients de ce codage, et sous quelles conditions il pourrait être envisageable d'utiliser ce codage pour la diffusion audio de programmes télévisés dans un futur proche.

Chapitre 1 : Le Son à la Télévision : Historique et État des lieux

Pour commencer, je vais faire un rapide tour d'horizon de ce qu'est la télévision, les innovations majeures qu'elle a connues, et je vais détailler les différents vecteurs de diffusion et les nouveaux modes de consommation de ce média.

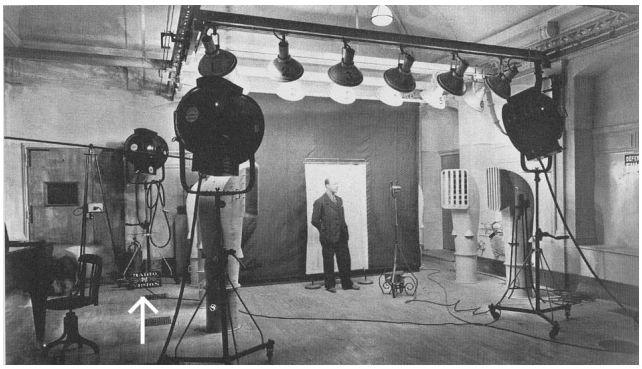
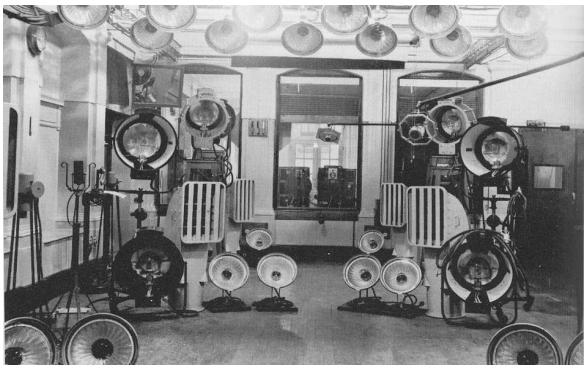
1. Historique de la Télévision Française

Dans les années 1880, plusieurs chercheurs pensent qu'on pourrait transmettre une image de télévision en la projetant sur une surface photosensible composée de points de sélénium, un matériau photo-électrique. Chaque point est transmis séquentiellement à un récepteur synchronisé à l'émetteur : c'est le principe de base de tout système de transmission d'images animées.

Le 14 Avril 1931, l'ingénieur français René Barthélémy réalise les premières transmissions d'images de 30 lignes de Montrouge à Malakoff, en banlieue parisienne. En Décembre 1932, il réalise un programme expérimental en noir et blanc d'une heure par semaine, nommé *Paris Télévision*. A cette époque, la France compte une centaine de récepteurs, essentiellement situés dans les services publics.

Le 26 Avril 1935, la première chaîne de télévision est créée, elle diffuse la première émission officielle de télévision française en noir et blanc, et en 60 lignes, depuis le ministère des PTT, 103 rue de Grenelle à Paris. En Novembre de la même année, le

premier émetteur d'ondes courtes est installé au sommet de la Tour Eiffel, la télévision passe alors en 180 lignes.

Figures 1 et 1bis : Studio au 103 rue de Grenelle, à Paris, en 1935
pour la première émission officielle de télévision française

En 1936, on recense environ deux milles récepteurs de télévision à travers le monde. Les Jeux de Berlin sont retransmis en direct et sont regardés par 150 000 téléspectateurs.

Dès 1937, des émissions sont diffusées tous les soirs entre 20h00 et 20h30, on compte alors une centaine de postes répartis chez les particuliers en France. Le son est monophonique puisque les téléviseurs sont équipés d'un seul haut-parleur. Au fur et à mesure, les programmes se développent.

En 1948, la télévision est émise en 819 lignes en France, tandis que tous les autres pays ont choisi une résolution de 625 lignes.

Les deux premières speakerines font leur apparition en mai 1949 : elles sont chargées de présenter les programmes. Le premier journal télévisé est diffusé le 29 Juin 1949, il est présenté par Pierre Sabbagh.

Figure 3 : Diffusion du premier journal télévisé en 1949

Figure 2 : 1er bulletin météo à la télévision

En 1950, on compte environ 3700 postes de télévision en France, ce chiffre est faible mais seulement 10% du territoire national est couvert. Deux ans plus tard, on en dénombrera près de 60 000. Le premier bulletin météorologique est diffusé en 1953. En 1958, environ un million de foyers français possèdent un téléviseur.

Figure 4 : Téléviseur à tube cathodique de 1954, avec un seul haut-parleur

La deuxième chaîne en noir et blanc voit le jour le 18 Avril 1964, avec une définition de 625 lignes. Cette chaîne sera émise en couleurs le 1er Octobre 1967 grâce à l'invention du SECAM (SÉquentiel Couleur À Mémoire), standard de codage de vidéo analogique en couleurs, à 625 lignes, créé par l'ingénieur français Henri-de-France. Les premières publicités apparaissent à la fin des années soixante. La couleur révolutionne la télévision. Le son est toujours monophonique.

Le 31 Décembre 1972, la troisième chaîne de télévision en couleurs, en 625 lignes, est lancée. Elle diffuse chaque soir des décrochages régionaux.

La première chaîne privée et payante naît en 1984 : Canal Plus.

En 1986, le son diffusé à la télévision devient stéréophonique.

Le Conseil Supérieur de l'Audiovisuel (CSA) naît le 12 Janvier 1989, en remplacement de la CNCL (Commission nationale de la communication et des libertés), il est alors composé de neuf membres élus pour six ans, par le Président de la République, le Président du Sénat et le Président de l'Assemblée Nationale. Le CSA est une autorité indépendante, qui a pour but de garantir l'exercice de la liberté dans le domaine de la communication audiovisuelle. Cet organisme deviendra aussi rapidement censeur.

Dès 1990, 50% des films et œuvres audiovisuelles diffusés doivent être d'origine française, et 60% doivent être d'origine communautaire.

A partir de Juillet 1991, la chaîne Antenne 2 (devenue France 2 en 1992) diffuse désormais vingt-quatre heures sur vingt-quatre.

En 1995, tous les téléviseurs vendus sont équipés d'un décodeur Nicam Stéréo, qui leur permet de décoder les programmes en stéréophonie. Le procédé Nicam encode l'audio à 32 kHz avec 14 bits compressés sur 10 bits (contre 44,1 kHz et 16 bits pour une qualité CD). Le Nicam disparaitra avec l'arrêt de la diffusion analogique.

En 2002, la première offre de télévision sur IP est proposée en France par l'opérateur Free, avec des chaînes gratuites et des chaînes payantes, vendues en bouquet ou à l'unité, sans durée d'engagement.

Figure 5 : Logo de la TNT en définition standard

Figure 6 : Logo de la TNT Haute Définition

Le 31 Mars 2005, la télévision numérique terrestre (TNT) est lancée. En France, la TNT exploite la norme DVB-T. La TNT en résolution d'image standard est diffusée en MPEG-2. Elle propose dès lors dix-huit chaînes gratuites. La télévision numérique est réceptionnée de façon hertzienne, avec une antenne râteau, comme en télévision analogique. La télévision numérique terrestre a permis une amélioration de la qualité de l'image, du son, et du graphique des sous-titrages.

Le 30 Octobre 2008, la TNT Haute Définition fait son apparition, avec quatre chaînes : TF1 HD, France 2 HD, M6 HD et Arte HD. Les chaînes en haute définition exploitent la norme H.264/MPEG-4 AVC. La diffusion en Haute Définition permet d'améliorer la qualité de l'image, qui a désormais une résolution de 1280 x 720 pixels, en format 16/9ème, et le son diffusé est alors un son multicanal 5.1.

Fin 2011, la diffusion analogique est totalement arrêtée en France.

Courant 2012, la TNT compte toujours dix-huit chaînes gratuites en définition standard : TF1, France 2, France 3 (nationale et locale), France 5, M6, Arte, D8, W9, TMC, NT1, NRJ12, La Chaîne Parlementaire – Public Sénat, France 4, BFMTV, iTélé, D17, Gulli, France Ô, ainsi que quatre chaînes en haute définition : TF1 HD, France 2 HD, M6 HD et Arte HD. Les programmes de Canal Plus en clair sont accessibles aussi en définition standard ou en haute définition.

Depuis le 12 Décembre 2012, six nouvelles chaînes gratuites en haute définition ont fait leur apparition sur la TNT : HD1, l'Équipe 21, 6ter, Numéro 23, RMC Découverte et Chérie 25, qui ont été déployées à Paris, à Marseille, en Aquitaine (Bordeaux, Bayonne) dans l'Yonne (Auxerre, Sens), à Troyes, en Bretagne (Rennes, Brest, Vannes), dans les Pays de la Loire (Angers, Le Mans, Tours) et dans le Centre. Les déploiements dans les autres régions françaises dureront jusqu'en Juin 2015.

Figure 7 : Les six nouvelles chaînes haute définition de la TNT

2. Écouter et Regarder la Télévision en 2013

En 2013, on peut désormais regarder la télévision grâce à différents vecteurs de diffusion : par la TNT, par le câble ou le satellite, par un réseau utilisant le protocole IP, et sur différents supports : sur le téléviseur évidemment, mais aussi sur des supports qu'on n'aurait jamais imaginés il y a vingt ans, comme par exemple l'ordinateur, une tablette ou un smartphone. Et depuis quelques années, on voit naître de nouveaux modes de consommation de ce média.

2.1. Les Vecteurs de Diffusion Télévisuelle

Aujourd'hui, la télévision est regardée par le biais de différents vecteurs de diffusion.

2.1.1. La Télévision Numérique Terrestre

La télévision numérique terrestre (ou TNT) est un mode de diffusion terrestre, dans lequel les signaux audio, vidéo et de données ont été numérisés, puis multiplexés, c'est-à-dire intégrés dans un flux unique, avant d'être modulés puis transportés jusqu'au téléspectateur en ondes électromagnétiques.

Les chaînes de la TNT sont réparties sur différents multiplex[3]. Un multiplex peut contenir jusqu'à six chaînes en définition standard ou trois chaînes en haute définition. En France, on compte huit multiplex, la figure 8 illustre la répartition des chaînes entre les différents multiplex. Le codage source est alors réalisé au niveau de la tête de réseau, juste avant le multiplexage, afin de réduire les ressources nécessaires à la transmission d'un programme, et garantir une qualité homogène des programmes en permanence. Le transport des multiplexes entre la tête de réseau (nationale ou régionale) et les sites de diffusion se fait par satellite ou via le réseau terrestre (faisceau hertzien, fibre optique).

Figure 8 : Répartition des chaînes entre les huit multiplexes

[3] Auparavant, en analogique, une seule chaîne de télévision était diffusée sur une fréquence ou un canal. Désormais, grâce à la technologie numérique, plusieurs chaînes peuvent être diffusées sur un même canal ou fréquence : le terme multiplex désigne donc un ensemble de chaînes numériques diffusées sur la même fréquence.

La télévision numérique terrestre exploite la norme DVB-T[4], avec un codage en MPEG-2 en définition standard et un codage MPEG-4 AVC pour les chaînes en haute définition. Cette norme impose une bande passante de 8 MHz, et un débit maximal de 25 Mbits/seconde. Elle exploite les bandes de fréquences III (174-230 MHz), IV (470-582 MHz) et V (582-862 Mhz). Les chaînes payantes sont souvent diffusées en MPEG-4 AVC, sauf les plages en clair qui sont diffusées en MPEG-2. La diffusion en haute définition utilise le format 1080i/50 Hz, avec un débit d'environ 8 Mbits/seconde.

A l'exception de la chaîne France 3, qui est codée en débit fixe pour permettre les décrochages régionaux, l'image de toutes les chaînes est codée en débit variable, afin de permettre, au programme le plus gourmand en débit (par exemple une compétition sportive ou un film d'action), d'utiliser davantage de débit à l'instant t, pour garantir une bonne qualité. En revanche, l'audio est toujours codé à débit fixe pour un programme (le débit peut éventuellement varier d'un programme à l'autre, bien que ce soit très rare).

La télévision numérique terrestre diffuse actuellement vingt-cinq chaînes nationales gratuites, ainsi qu'une quarantaine de chaines locales au total, accessibles sans abonnement (les chaînes locales disponibles dépendent de la région où l'on se situe).

Figure 9 : Les vingt-cinq chaînes gratuites de la TNT

[4] La norme DVB-T est l'application de la norme de diffusion de la télévision numérique par liaisons hertziennes terrestres (Digital Video Broadcasting - Terrestrial, en français Diffusion vidéo numérique terrestre).

Quatre chaînes gratuites sont diffusées en définition standard et en haute définition, il s'agit de TF1, France 2, M6 et Arte. Il existe aussi onze chaînes nationales payantes.

Figure 10 : quelques chaînes locales diffusées sur la TNT

La réception de la télévision numérique terrestre se fait toujours grâce aux antennes râteaux existantes, individuelles ou collectives, qui étaient utilisées auparavant pour la réception de la télévision analogique. Le signal peut alors être décodé de différentes façons : soit à partir d'un décodeur TNT relié en Péritel à un téléviseur analogique (tube cathodique) ou à un téléviseur numérique non muni d'un décodeur intégré, soit directement par un décodeur intégré dans un téléviseur numérique (ce qui est le cas pour tous les téléviseurs fabriqués depuis 2008), soit avec une antenne intérieure en réception en mode portable.

Sur les téléviseurs récents, le téléspectateur peut choisir de visionner la chaîne en définition standard ou en haute définition. Il peut aussi choisir son canal audio (version originale, version multilingue, audio-description) et le sous-titrage (sans sous-titrage, sous-titrage multilingues ou sous-titrage pour sourds et malentendants).

La télévision numérique terrestre couvre seulement 85% de la population française.

2.1.2. La Télévision par Câble

La télévision numérique par câble est née en 1985, elle diffuse des programmes télévisés par l'intermédiaire d'un réseau câblé. Ce réseau câblé est constitué de trois parties : la station de tête, le réseau de télédistribution et le terminal. La station de tête combine des antennes terrestres qui captent les chaînes nationales et locales de la TNT, des antennes paraboliques qui captent les chaînes numériques diffusées par des satellites géostationnaires, des canaux radiophoniques, les signaux de télécommunication et multimédias IPTV qui permettent l'accès à la vidéo à la demande, ainsi qu'à Internet et à la téléphonie. Une fois captés, ces signaux sont traités : ils sont filtrés et convertis en d'autres fréquences, démodulés, décodés, décryptés voire transcodés afin de les adapter aux normes en vigueur, réordonnés, multiplexés et adaptés au contrôle d'accès spécifique du réseau câblé. Tous ces signaux composent un plan de fréquence qui exploitent les bandes de fréquences de la télévision : VHF bandes I (47-68 MHz) et III (174-223 MHz), et UHF bandes IV (470-582 MHz) et V (582-862 MHz). Les signaux sont transportés de la tête de réseau à la prise d'abonné par un câble coaxial avec des amplificateurs de ligne ou par fibre optique monomode, en passant par des amplificateurs de distribution.

Dans les grandes villes, les réseaux câblés sont désormais hybrides bidirectionnels, avec un transport à la fois sur câble coaxial et sur fibre optique. La diffusion par câble utilisant la norme DVB-C[5], il faut alors posséder un téléviseur équipé d'un récepteur compatible ou utiliser un récepteur externe.

En 2006, les deux câblo-opérateurs UPC-Noos et Numericable ont fusionné, et aujourd'hui Numericable est donc le seul opérateur de diffusion câblée en France. L'inconvénient majeur de cette diffusion câblée est le coût élevé de cette technologie.

Figure 11 : Logo de Numericable, opérateur câblé

[5] La norme DVB-C est l'application de la norme Digital Video Broadcasting aux transmissions par câble. La bande disponible de fréquences est réduite à 8 MHz, le rapport signal à bruit est correct, les perturbations observées sont dues à une mauvaise adaptation de la prise utilisateur.

2.1.3. La Télévision par Satellite

La télévision par Satellite consiste à émettre des programmes radiophoniques, télévisés, analogiques ou numériques, payants ou gratuits, depuis un satellite en orbite géostationnaire. Le principal avantage est la zone de couverture, supérieure à 99,5% du territoire français (les zones non couvertes se situent sur certaines faces abruptes des massifs montagneux ou dans certaines zones urbaines à cause de grandes tours). C'est aussi le moyen de diffusion le moins onéreux, par rapport à la TNT et au câble. La diffusion par satellite exploite la norme DVB-S[6], avec un codage audio et vidéo en MPEG-2 pour les programmes en définition standard et en MPEG-4 AVC pour la vidéo en haute définition. Les diffuseurs français et européens utilisent les satellites Hot Bird et Atlantic Bird 3 d'Eutelsat et Astra 19,2°E de SES S.A.. En diffusion numérique satellite, la bande de fréquences disponible est plus large : 36 MHz.

Pour réceptionner la télévision par satellite, il faut une antenne parabolique équipée d'une tête universelle, un câble coaxial, un démodulateur compatible DVB-S et un câble, Péritel ou HDMI, pour relier le décodeur au téléviseur. Des fabricants commercialisent désormais des téléviseurs avec tuners DVB intégrés mixtes (DVB-T et DVB-S), afin de permettre aux téléspectateurs non couverts par la TNT de recevoir les chaînes de télévision gratuites sans ajouter de récepteur externe. Les décodeurs compatibles DVB-S doivent être équipés d'un lecteur de cartes pour pouvoir recevoir des chaînes payantes, moyennant un abonnement.

TNTSAT et FRANSAT proposent toutes les chaînes gratuites de la TNT, y compris les chaînes haute définition, ainsi que des chaînes locales gratuites (et notamment les vingt-quatre éditions régionales de France 3), quelques chaînes thématiques, des radios, et des bouquets disponibles sur abonnement, comme par exemple Canal Plus, CanalSat, SFR et la TV d'Orange. Le bouquet CanalSat était initialement un bouquet de télévision par satellite, lancé en 1992 en analogique, puis en 1996 en numérique.

[6] La norme DVB-S est l'application de la norme Digital Video Broadcasting pour les transmissions par satellite.

Dans les années futures, la parabole permettra une liaison « retour » vers le satellite qu'elle capte, afin de proposer des services de vidéo à la demande aux clients.

2.1.4. La Télévision sur IP (Internet Protocol)

Un nouveau vecteur de diffusion est apparu en 2002 : la télévision numérique sur IP ou IPTV, diffusée sur un réseau qui utilise le protocole IP. Le flux télévisé est transporté par la ligne téléphonique, avec un mode de transport basé sur le protocole Internet (IP) grâce à la technologie ADSL[7]. Le téléspectateur reçoit la télévision grâce au boîtier appelé « set-top boxe », fourni par les fournisseurs d'accès à Internet (FAI), moyennant un abonnement, souvent sous la forme d'une offre globale « triple play » comprenant l'accès à l'Internet ADSL haut débit illimité, les appels téléphoniques vers les numéros fixes en illimité et la télévision avec plus de cent chaînes gratuites. La télévision numérique sur IP regroupe différents services : la télévision en direct, la vidéo à la demande, la télévision de rattrapage, des jeux à la demande (Game on Demand ou GoD). Le boîtier permet de décoder et de décrypter les flux télévisés et de vidéo à la demande, afin de les lire sur le téléviseur.

La technologie IP propose davantage de contenus et de fonctionnalités. Contrairement à un réseau de télévision classique ou satellite, où tous les contenus sont constamment offerts à l'utilisateur qui les sélectionne ensuite dans son décodeur, la télévision sur IP fonctionne différemment : tous les contenus restent sur le réseau, et seuls ceux qui sont sélectionnés par l'utilisateur lui sont alors envoyés, la bande-passante nécessaire est donc réduite. Les chaînes de télévision sont diffusées en multicast : c'est une diffusion multi-points, d'un émetteur unique vers un groupe de récepteurs, tandis que

[7] ADSL : Assymetric Digital Subscriber Line, ou liaison numérique asymétrique, technique de communication numérique massivement utilisée par les Fournisseurs d'Accès à Internet pour les connexions dites à « haut-débit ». Cette liaison est dite asymétrique car son débit de données montant (upload) est entre cinq et vingt fois inférieur au débit de données descendant (download).

les vidéos à la demande sont envoyées en unicast, c'est-à-dire en diffusion point à point, la sélection du programme se faisant au niveau du DSLAM (Digital subscriber line access multiplexer, ou Multiplexeur d'accès DSL), grâce au caractère bidirectionnel de la liaison ADSL. Le DSLAM est un boîtier situé au nœud de raccordement des abonnés, dans le central téléphonique, qui fait la liaison entre les lignes d'abonnés et le réseau de l'opérateur auquel il appartient. Un DSLAM peut gérer 500 à 1000 lignes. La principale contrainte de la télévision numérique sur IP est d'avoir un débit suffisant, qui admette à la fois la lecture d'un flux d'images vidéo de façon fluide (d'un débit allant de 1,5 à 6 Mbits/seconde selon la résolution vidéo), et une navigation sur internet. Pour que le débit soit suffisant, les abonnés doivent résider à moins de quatre kilomètres du nœud de raccordement des abonnés.

L'accès Internet à haut débit promettait initialement un débit pouvant atteindre 8 Mbits/seconde, puis 20 Mbits/seconde dès 2004 avec l'ADSL 2+. Le développement de l'accès à Internet à Très Haut Débit (allant de 30 à 100 Mbits/seconde) grâce à la fibre optique, permet désormais d'envisager de nouvelles innovations, comme par exemple davantage de chaînes en haute définition, des chaînes en 3D, de la vidéo à la demande en haute définition, etc.

La télévision sur IP a révolutionné les modes de consommation de ce média, en proposant des services interactifs, comme par exemple un guide des programmes, ou encore la fonctionnalité Picture in Picture (PiP), qui permet de visionner dans un petit encart une autre chaîne ou encore de choisir un angle caméra sur un match de football par exemple.

L'opérateur d'accès à Internet Free a été le premier en France à proposer des offres de télévision numérique sur IP, il a ensuite été rejoint par Orange (France Telecom), SFR (anciennement 9 Telecom), Bouygues Telecom, Alice et plus récemment Darty.

Selon Médiamétrie, au troisième trimestre 2012, plus de 70% des foyers français disposent d'un abonnement Internet à haut débit, et 69% de la population équipée d'un

téléviseur au moins et résidant en France reçoit la télévision par le câble, le satellite ou l'ADSL[8]. 37,1% des foyers français reçoivent la télévision par l'ADSL.

2.1.5. Répartition des Vecteurs de Diffusion dans les Foyers Français

Le Conseil Supérieur de l'audiovisuel a publié les résultats d'une étude de l'Observatoire de l'équipement des foyers pour la réception de la télévision numérique : au second semestre 2011, 26,6 millions des foyers français, soit 99% des foyers équipés d'au moins un téléviseur, avaient accès à la télévision numérique, dont 50% qui n'utilisaient plus que la télévision numérique terrestre.

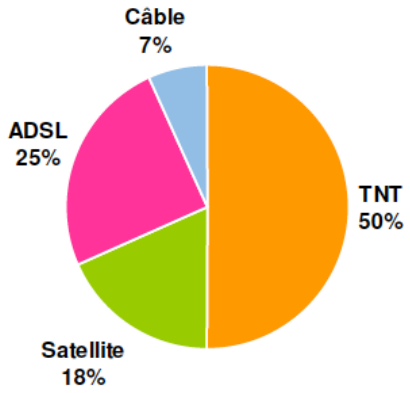


Figure 12 : Répartition des différents vecteurs de diffusion de télévision numérique dans les foyers français

2.2. Les nouveaux modes de consommation de la télévision

Depuis les débuts de la télévision dans les années 1930, les modes de consommation des flux télévisés ont beaucoup évolué, notamment avec la multiplication des chaînes gratuites et payantes et les nouveaux services proposés.

2.2.1. La télévision de rattrapage

La télévision de rattrapage, encore appelée replay ou catch-up tv, est un service qui propose de revisionner gratuitement un programme, souvent pendant les sept jours suivants la diffusion (bien que certains programmes soient en ligne jusqu'à trente jours après la diffusion). Ce service, d'abord proposé par la télévision numérique sur IP,

[8] Selon un communiqué de presse de Médiamétrie, publié le 19/03/2013.

concerne une large gamme de programmes (journaux télévisés, talk-shows, magazines, séries et téléfilms, jeux, divertissements, documentaires), à l'exception des films cinéma. En Octobre 2012, on dénombrait environ 59% des programmes des chaînes disponibles en replay[9].

Désormais, l'audimat est mesuré en tenant compte des visionnages en catch-up tv, et on appelle donc ces mesures « audiences consolidées ». En un an, 810 000 téléspectateurs de plus regardent régulièrement des programmes en catch-up TV sur leur téléviseur. En Mars 2013, 17,9% des Français (soit un peu moins d'un Français sur 5) a regardé au moins un programme en télévision de rattrapage au cours du mois[10].

2.2.2. LA VIDÉO À LA DEMANDE

Les fournisseurs d'accès à Internet et les opérateurs câblés proposent aussi un service de vidéo à la demande (VàD en français ou VoD en anglais) : il s'agit d'une technique de diffusion interactive de contenus vidéos numériques, moyennant l'achat à l'unité d'un programme pour une durée limitée ou un abonnement supplémentaire à ce service (on parle alors de Vidéo à la demande avec abonnement, abrégée en VàDA en français ou SVoD en anglais). Le téléspectateur loue alors un programme, souvent un film cinéma, pour une durée déterminée (généralement pendant vingt-quatre heures), durée pendant laquelle il pourra visionner en streaming ce programme, disposant des fonctions « pause », « avance rapide » ou « retour rapide ». Ce service est apparu au début des années 2000, la diffusion se fait en mode unicast, c'est-à-dire point à point, puisque l'utilisateur choisit le programme et le moment où il veut le regarder, mais cette technologie requiert beaucoup de ressources réseau. 70% des programmes disponibles en

[9] D'après une étude du CNC nommée « Baromètre de la télévision de rattrapage – octobre 2012 », publiée le 11/12/2012.

[10] Selon un communiqué de presse de Médiamétrie intitulé « Davantage d'écrans pour plus de télévision en direct ou en rattrapage », daté du 05/04/2013.

vidéo à la demande sont des films cinéma, 16% des films pour adultes, 9% sont des fictions télévisées et les 5% restant regroupent tous les autres types de programmes[11].

2.2.3. L'Hybrid Broadband Broadcast TV (HbbTV)

Un consortium, formé par fabricants et des diffuseurs européens, a créé l' « Hybrid Broadband Broadcast TV » ou HbbTV en 2010, qui est à la fois un standard industriel et une harmonisation de la diffusion télévisuelle et de l'accès Internet dans les téléviseurs connectés et les boîtiers décodeurs. Ce standard européen de diffusion est dédié à la télévision connectée, et propose des services associés ou non aux programmes de télévision en cours de diffusion. Il utilise les standards de l'Internet comme HTML, JavaScript, CSS et un codec vidéo H264.

Figure 13 : Logo de l'HbbTV

Ce standard permet aux chaînes de télévision d'enrichir leur flux traditionnel avec des contenus additionnels qui accompagnent leurs programmes et d'homogénéiser ce service interactif, qui ne dépend donc plus du fabricant du décodeur. Il existe deux modes : un mode appelé « broadcast » et un mode dit « broadband ». Le mode « broadcast » ne nécessite pas de connexion à Internet, une application se télécharge en même temps que le programme télévisé est regardé, que ce soit en TNT, par le câble ou le satellite. Parmi les services proposés, on trouve notamment des fiches contextuelles. Si le récepteur est connecté à Internet, on est alors en mode « broadband » : les chaînes peuvent alors s'enrichir des contenus en ligne, édités par elles-mêmes. Le téléspectateur peut ainsi accéder à davantage de services, tels que des vidéos à la demande, des programmes en replay, des émissions enrichies, des votes en direct, un guide électronique des programmes (EPG), un choix de la caméra sur un programme tourné en multi-angles de prise de vue, la météo, etc. Plusieurs pays européens utilisent ce standard ; en France, les chaînes ARTE , France 2, France 3, France 4, France 5, France Ô, NRJ12, TF1 et iTélé proposent des services de HbbTV.

[11] Selon une étude de GFK et NPA Conseil, datée de novembre 2011.

2.3. Les Supports de Visionnage de la Télévision

Aujourd'hui, il existe différents types de téléviseurs. On trouve encore des téléviseurs à tube cathodique dans certains foyers français, malgré l'arrêt total de la diffusion analogique hertzienne. Pour recevoir la télévision numérique terrestre, ces personnes possèdent donc un décodeur externe, relié en Péritel au téléviseur.

Au début des années 2000 sont apparus les écrans plats, qui se sont petits à petits démocratisés et que l'on retrouve dans la plupart des foyers français aujourd'hui. Il existe différentes technologies de téléviseurs à écrans plats : les écrans à cristaux liquide (LCD), les écrans à plasma, les écrans à diode électroluminescente (LED) et depuis 2009 des écrans 3D. Tous les téléviseurs intègrent un décodeur TNT depuis 2008. En 2005, 8% de foyers français possédaient un écran plat, en 2012 ils étaient 88%. De plus, la taille moyenne des écrans vendus en France augmente chaque année, passant de 25,5 pouces en 2006 à 33,3 pouces en 2012.[12]

Grâce aux boîtiers appelés « set-top box » (ou boîtier décodeur télévisé) et aux offres « triple play » des fournisseurs d'accès à Internet, les téléspectateurs ont découvert le principe de la télévision connectée : guide interactif des programmes, télévision de rattrapage (TVR, aussi appelée replay ou catch-up tv), vidéo à la demande (VàD et SVoD), les podcasts radio, la navigation sur le web, les jeux, la fonction media center pour les boîtiers qui intègrent un disque dur et permettent l'enregistrement de contenus vidéos (ces boîtiers sont nommés « Personal Video Recorder » (PVR) en anglais ou « numériscope » en français (ou encore magnétoscope numérique)), ainsi qu'un accès à des ressources externes grâce au port USB. Depuis quelques années, les téléviseurs dits « SMART TV » ou « téléviseur connectable » sont apparus sur le marché : il s'agit de téléviseurs que l'on peut connecter à Internet, et qui essaient de concurrencer les offres des Fournisseurs d'Accès à Internet, en proposant des contenus de type guide interactif des programmes ou télévision de rattrapage. Pour bénéficier des services de télévision connectée si notre écran n'est pas compatible, il existe aussi des boîtiers dédiés, comme

[12] Selon une étude de GFK – Médiamétrie, publiée en février 2013.

par exemple l'Apple TV, qui se connectent à des environnements propriétaires, ou des boîtiers hybrides pour accéder aux services HbbTV. Les fabricants Sony, Samsung et LG proposent des téléviseurs connectables, néanmoins les téléspectateurs plébiscitent encore les offres des FAI, qu'ils trouvent plus complètes et plus ergonomiques : ils veulent regarder des contenus vidéos et préfèrent utiliser leur ordinateur, tablette ou smartphone pour faire des recherches de programmes ou pratiquer la social tv[13]. Parmi les utilisateurs des services de télévision connectée, 14% d'entre eux possèdent un téléviseur connectable, 37% profitent des services via leur box ADSL et leur offre « triple play », et 20% via une console de salon connectée.[14] Au troisième trimestre 2012, on dénombre 13% des foyers français équipés de « SMART TV », mais seulement 58% d'entre eux l'ont connectée à Internet.

On peut aussi regarder la télévision depuis un ordinateur, une tablette ou un smartphone. L'ordinateur est le premier écran utilisé pour la télévision de rattrapage, tandis qu'en mars dernier, 5,2% des détenteurs de smartphone et 13,5% des équipés en tablette numérique ont pratiqué la télévision de rattrapage sur ces supports.

Le téléviseur reste l'écran privilégié par le public pour consommer des contenus télévisuels, en effet en mars 2013, 98% des individus de 15 ans et plus ont regardé la télévision en direct ou en différé sur un téléviseur, et parmi eux 19,4% ont aussi regardé un programme en live ou en catch up sur ordinateur, 6,5% sur un smartphone et 3,3% sur

[13] L'expression « Social TV » désigne l'interaction entre le contenu des programmes télévisés (en direct ou en rattrapage) et les réseaux sociaux. Ceci est directement possible avec les téléviseurs connectés, néanmoins les téléspectateurs préfèrent souvent leur smartphone ou leur tablette pour échanger autour des programmes sur les réseaux sociaux.

[14] Selon une étude de GFK – Médiamétrie, publiée en février 2013.

une tablette, soit 23,4% des téléspectateurs qui visionnent des contenus télévisés sur d'autres écrans[15].

En 2012, les Français ont passé en moyenne 3 heures 50 minutes chaque jour à regarder la télévision sur un téléviseur, soit trois minutes de plus qu'en 2011[16].

2.4. Les Supports d'Écoute de la Télévision

Tout comme l'image, il existe différents systèmes pour écouter la télévision : les haut-parleurs du téléviseur, un système stéréophonique externe (deux enceintes externes actives ou passives et connectées à un amplificateur), ou encore un système home-cinéma, une barre de son ou plus rarement au casque (avec ou sans fil).

L'écoute avec les haut-parleurs du téléviseur est la plus courante, cette solution évite d'investir dans un équipement audio mais la qualité est très moyenne, voire médiocre.

Un système home-cinéma audio est composé d'un système de décodage audio multicanal (Stéréo, Dolby Pro Logic, Dolby Pro Logic 2, Dolby Digital et Dolby Digital Plus, LC Concept, DTS, SDDS, Dolby TrueHD et DTS HD au minimum), une amplification multi-voies et plusieurs enceintes. Le décodeur et l'amplificateur sont souvent inclus dans un même appareil. Ce système est au minimum un système 5.1, composé de cinq enceintes et d'un caisson de basse. On trouve parfois deux enceintes arrières en plus ou deux enceintes latérales. Un home-cinéma multicanal présente l'avantage de pouvoir écouter les programmes en multicanal diffusés sur la TNT en haute définition. Mais il est difficile à mettre en place chez soi, puisqu'il nécessite un

[15] Selon un communiqué de presse de Médiamétrie intitulé « Davantage d'écrans pour plus de télévision en direct ou en rattrapage », daté du 05/04/2013.

[16] D'après Médiamétrie – Global TV, enquête réalisée en octobre et novembre 2012.

aménagement du salon, avec des câbles qui traînent, et est particulièrement laborieux à intégrer dans les grandes pièces à vivre dites « ouvertes ».

L'avenir des systèmes d'écoute de la télévision est probablement la barre de son, qui regroupe des enceintes et l'amplificateur audio en un seul boîtier long de 90 cm environ, que l'on place sous l'écran. Il faut juste ajouter un caisson de basse. La barre de son utilise les réflexions des murs du salon pour recréer les canaux arrières. L'installation est donc simplifiée et aucun câble audio ne traverse la pièce, néanmoins ce système est onéreux et nécessite une géométrie de pièce particulière. Ce type d'installation n'est pas encore répandu, bien que prometteur.

Environ 14,2% des foyers français sont équipés d'un système home-cinéma[17].

Avec l'évolution des modes de consommation de la télévision, qui devient désormais mobile, on peut alors écouter la télévision au casque avec fil, sur un smartphone, une tablette ou un ordinateur, ou bien même dans son salon avec un casque avec ou sans fil, afin de ne pas déranger ses colocataires.

[17] Selon une étude de GFK – Médiamétrie du 2ème trimestre 2011.

CHAPITRE 2 : NOTIONS DE BASE DE LA PERCEPTION AUDITIVE HUMAINE

Avant d'étudier le codage MPEG Surround, il est important de rappeler les notions de base de la perception auditive humaine, puisque le MPEG Surround utilise des propriétés psycho-acoustiques essentielles pour que l'auditeur puisse localiser une source sonore.

1. L'OREILLE HUMAINE

1.1. ANATOMIE DE L'OREILLE

L'oreille est un organe composé de trois parties : l'oreille externe, l'oreille moyenne et l'oreille interne. L'oreille externe, composée du pavillon et du conduit auditif, est chargée de transmettre la vibration acoustique au tympan. L'oreille moyenne, qui regroupe le tympan, les trois osselets (le marteau, l'enclume et l'étrier), est un transducteur, qui transforme la vibration acoustique en une vibration mécanique. Enfin, l'oreille interne, qui rassemble la trompe d'Eustache, le vestibule, la cochlée et le nerf auditif, transforme la vibration mécanique en un influx nerveux, qui est ensuite véhiculé par le nerf auditif jusqu'au cortex auditif du cerveau, où il sera décodé.

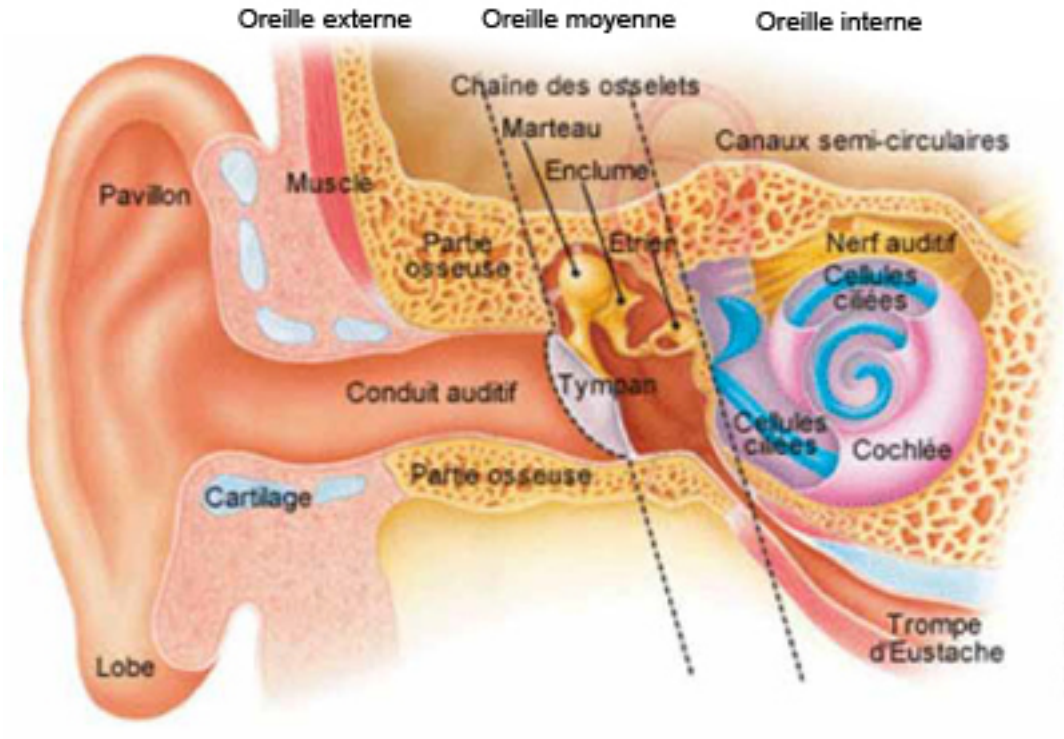


Figure 14 : Schéma d'une oreille

1.2. Seuil Absolu d'Audition et Sensibilité de l'Oreille

La sonie désigne le niveau sonore perçu, elle dépend principalement du niveau de pression sonore mais aussi de la fréquence et de la durée du son. Un son pur d'une fréquence de 1 kHz et d'un niveau de pression sonore de 40 dB a une sonie d'un sone. Pour un son donné, le niveau d'isosonie est mesuré en phones, il désigne le niveau de pression acoustique d'un son pur de 1 kHz de la même sonie.

En théorie, l'oreille humaine perçoit tous les sons dont les fréquences sont comprises entre 20 Hz et 20 kHz, dans une gamme de niveaux sonores de 0 dB à 120 dB (120 dB étant le seuil de douleur). La figure 15 montre les courbes d'isosonie de Fletcher et Munson : on remarque que l'oreille est très sensible dans l'intervalle entre 2 et 5 kHz, tandis que sa sensibilité est moindre dans les extrêmes graves et aigus.

Sur cette même figure, la courbe en pointillé correspond au seuil absolu d'audition, qui est le niveau minimal de pression acoustique qu'il faut appliquer à un son pur pour qu'il soit audible par un être humain dans un environnement silencieux. L'oreille est donc un système non-linéaire à seuillage adaptatif.

Évidemment, les courbes de sensibilité de l'oreille sont propres à chaque individu, elles dépendent de l'âge, mais aussi des caractéristiques physiologiques de chacun, et notamment des éventuels traumatismes subis (une longue exposition à des sons très forts peut altérer de façon irréversible l'audition, et ce quelque soit l'âge).

Une des premières étapes des codeurs perceptifs consiste donc à éliminer les signaux dont l'amplitude se situe en-dessous du seuil de perception.

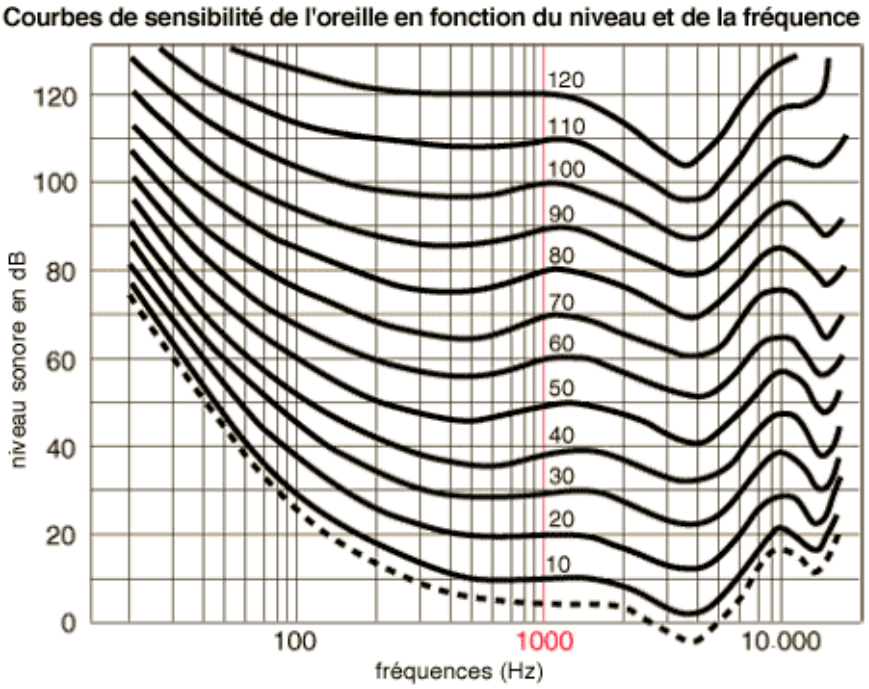


Figure 15 : Courbes d'isosonie de Fletcher et Munson (et seuil absolu d'audition en pointillé)

Des expériences menées par Harvey Fletcher et Eberhard Zwicker au XXème siècle ont montré que la membrane basiliaire, situé dans l'oreille interne, subdivise le spectre sonore audible par l'être humain en vingt-cinq sous bandes, définies par leurs fréquences centrales et leurs largeurs. Ces bandes sont appelées bandes critiques.

Afin de se rapprocher des caractéristiques de l'audition humaine, beaucoup de systèmes de codages perceptuels subdivisent le spectre en un certain nombre de sous-bandes de fréquences, souvent en trente-deux bandes de même largeur.

2. Effets de Masque

La courbe du seuil absolu d'audition n'est valable que pour un son pur écouté dans une ambiance calme. Lorsque les sons sont multiples, le seuil d'audition est modifié en permanence, à cause des phénomènes de masquage.

L'oreille humaine est incapable de distinguer deux sons émis à des fréquences proches ainsi que dans un temps réduit : on parle alors de masquage ou d'effet de masque.

Les masquages peuvent donc être fréquentiels et/ou temporels. Les codeurs perceptuels exploitent les phénomènes de masquage pour ne coder que les signaux seront audibles.

2.1. Masquage Fréquentiel

Le phénomène de masquage fréquentiel ou masquage simultané intervient lorsqu'un son fort, appelé son masquant, empêche la perception d'un son de niveau sonore plus faible, mais tout à fait audible s'il était seul, et ayant une fréquence voisine de celle du son fort. Un son masqué peut être un son pur ou un son occupant une bande de fréquences très étroite. Le seuil d'audition est alors altéré et remonté dans une bande adjacente à la fréquence centrale du son masquant, comme le montre la figure 16. Évidemment, lorsque plusieurs sons sont émis simultanément, il peut y avoir plusieurs sons masquants et plusieurs sons masqués. Il est alors inutile de coder ces derniers, puisqu'ils ne seront pas perçus. Le seuil d'audition varie donc à chaque instant, en fonction de la composition du signal.

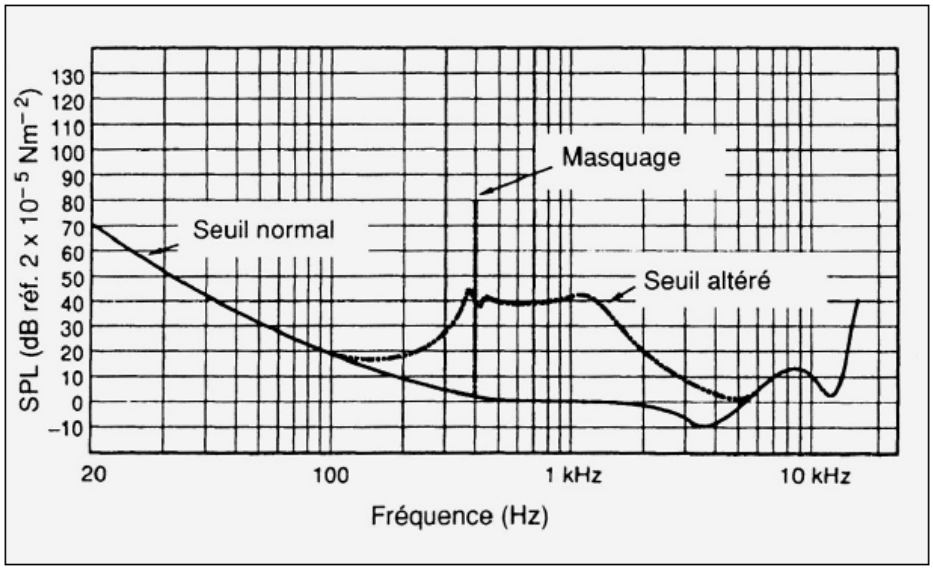


Figure 16 : Masquage fréquentiel par un son pur

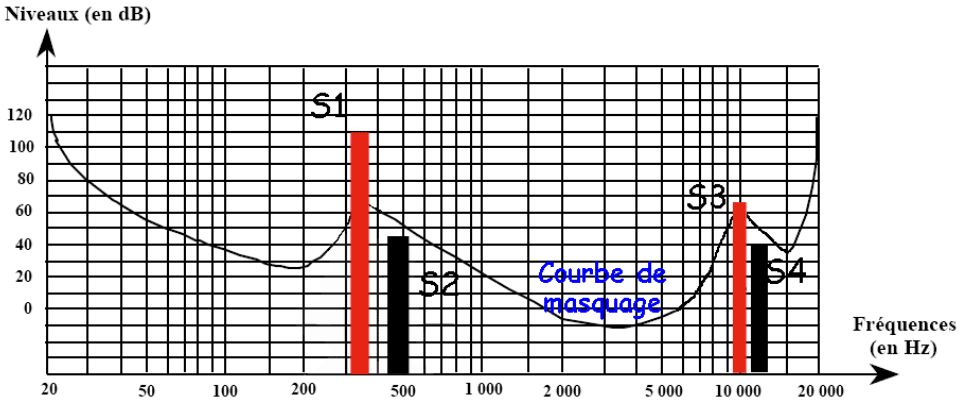


Figure 17 : Seuil d'audition altéré par deux sons forts

La figure 17 illustre ce phénomène de masquage fréquentiel : les sons S2 et S4 seraient parfaitement audibles s'ils étaient seuls. Mais les sons S1 et S3 sont beaucoup plus forts et deviennent masquants : ils modifient donc le seuil absolu d'audition autour de leur fréquence et masquent les sons S2 et S4.

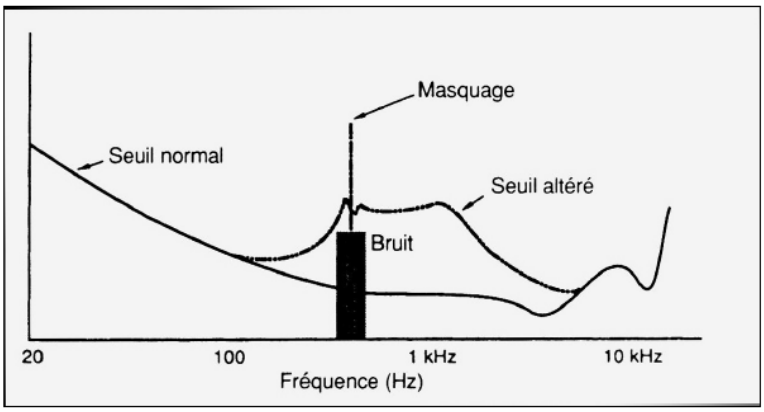


Figure 18 : Bruit masqué par un son pur

Dans la sous-bande où est inclus le son pur masquant et fort, on admet un bruit de quantification plus élevé, et donc une quantification moindre pour cette sous-bande car ce bruit sera masqué par le son lui-même. De plus, plus la largeur de la sous-bande est étroite, ce qui engendre un nombre élevé de sous-bandes et des algorithmes de codages plus complexes, plus le niveau de bruit admissible est élevé car il sera masqué. Ceci permet une réduction de débit supplémentaire.

Le masquage est d'autant plus prononcé lorsque le niveau du son masquant est fort, il est maximal lorsque le son masqué a une fréquence très proche du son masquant. De plus, les basses fréquences sont beaucoup plus masquantes. Les fréquences multiples entiers de la fréquence du son masquant observent aussi des dégradations. En revanche, deux sons de même niveau sonore ne peuvent pas se masquer.

2.2. MASQUAGE TEMPOREL

Lorsque deux sons sont émis successivement de façon rapprochée, il peut y avoir un masquage temporel, qui peut être un pré-masquage et/ou un post-masquage.

Un son de forte intensité modifie le seuil absolu d'audition, et peut masquer un son moins fort qui le précède juste : il s'agit de pré-masquage, qui a lieu quelques millisecondes avant le début du son masquant. Ce phénomène s'explique par l'inertie de l'oreille.

En outre, un son de forte intensité masque aussi des sons moins forts qui le suivraient : en effet, après un son de forte intensité, l'oreille ne peut plus percevoir des

sons plus faibles pendant quelques centaines de millisecondes, comme le montre la figure 19.

Ainsi deux sons émis de façon très rapprochée ou ayant une fréquence proche peuvent être tous les deux audibles, à condition que l'un ne masque pas l'autre.

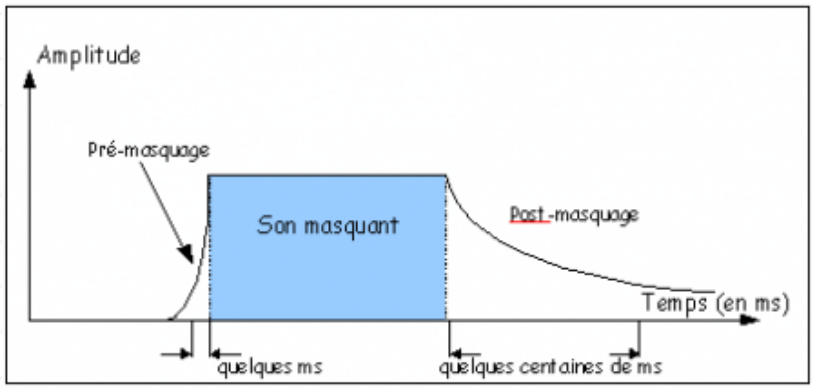


Figure 19 : Masquage temporel

3. Perception Auditive de la Spatialisation

L'être humain, comme la majorité des vertébrés, possède deux oreilles disposées à égale hauteur de chaque côté de la tête. Plusieurs indices permettent au cerveau d'analyser l'espace sonore environnant.

3.1. Indice de Différence Interaurale de Niveau (ILD) et indice de Différence Interaurale de Temps (ITD)

Afin de pouvoir définir la provenance d'un son, il existe deux indices : l'indice de différence interaurale de niveau (ILD, Interaural Level Difference) et l'indice de différence interaurale de temps (ITD, Interaural Time Difference).

L'indice de différence interaurale de niveau (ILD) est la différence d'intensité sonore entre un signal arrivant à l'oreille la plus exposée à la source et le signal arrivant à l'oreille opposée : en effet, si une source sonore est plus proche de l'oreille gauche, elle

arrivera avec un niveau plus élevé à l'oreille gauche qu'à l'oreille droite. Ce phénomène sera d'autant plus prononcé si la fréquence de la source sonore est élevée car la tête va d'autant plus agir comme un panneau.

L'indice de différence interaurale de temps (ITD) est la différence de temps d'arrivée d'un son entre les deux oreilles : par exemple, la même source sonore plus proche de l'oreille gauche que de l'oreille droite arrivera en premier à l'oreille gauche.

Ces deux indices permettent de bien discerner la gauche de la droite, en azimuth. Néanmoins, il existe une zone nommée cône de confusion où ces deux indices sont identiques pour différents positionnements de la source, en particulier en élévation : la perception de la spatialisation est alors floue.

3.2. Fonction de Transfert de la Tête (HRTF)

Il existe donc un troisième indice appelé monaural, indice spectral qui met en évidence le filtrage introduit par le corps humain : la tête, le pavillon de l'oreille, les cheveux, les épaules, le torse. C'est le seul indice qui diffère dans le cône de confusion. Cet indice peut s'exprimer dans le domaine temporel, nous parlerons de HRIRs (Head Related Impulse Response) ou réponse impulsionnelle due à la tête, ou peut s'exprimer dans le domaine fréquentiel, nous parlerons alors de HRTF (Head Related Transfer Function) ou fonction de transfert de la tête. Par définition, chaque être humain possède sa propre fonction de transfert de la tête.

La taille de la tête provoque un important filtrage dans les hautes fréquences, le pavillon de l'oreille entraîne des réflexions et des résonances qui modifient le spectre capté par le tympan, selon l'angle d'incidence du son. Les réflexions dues aux épaules et au corps bouleversent aussi le spectre. Toutes ces caractéristiques, propres à chaque être humain, conduisent à une unique fonction de transfert de la tête, pour chaque position et angle d'incidence, en azimuth et en élévation.

Quand la source sonore se situe à l'arrière, les hautes fréquences sont très filtrées. Afin de percevoir la latéralisation dans un espace, ce sont les différences de fonctions de transfert de la tête entre les deux oreilles qui aide le cerveau à localiser des sources, ainsi que la différence interaurale de temps.

De plus, les pavillons sont très différents d'un être humain à un autre, ce qui rend quasiment impossible une généralisation des fonctions de transfert de la tête. Le pavillon de l'oreille a une forme asymétrique, et procède comme un filtre acoustique de la source, en fonction de sa localisation. Ce filtrage agit essentiellement dans la bande de fréquences 6 à 10 kHz, et permet à l'auditeur de différencier l'avant de l'arrière, et de localiser la source en élévation.

Des personnes ont déjà testé une écoute avec des écouteurs (leurs pavillons n'interagissent donc pas) et des fonctions de transfert de la tête appartenant à un autre individu : leur capacité à localiser des sources sonores en est alors réduite, ce qui démontre qu'il est primordial de tenir compte des réflexions induites par ses propres pavillons.

De plus, la résonance du conche, partie de l'oreille qui mène au conduit auditif, est responsable de la création d'une impression d'externalisation, c'est-à-dire la sensation que le son provient de l'extérieur de la tête, contrairement à une écoute traditionnelle au casque.

Plusieurs centres de recherches ont réussi à mettre en évidence qu'avec un petit nombre de fonctions de transfert différentes, on réussissait à représenter la majorité de la population, néanmoins les meilleures expériences en binaural sont celles réalisées avec des fonctions de transfert de la tête personnalisées.

3.3. DEGRÉ DE COHÉRENCE

Le degré de cohérence interaurale (ICC) entre les signaux est un autre indice qui permet au cerveau d'analyser l'espace sonore qui l'entoure. En effet, une source sonore subit de nombreuses réflexions avant de parvenir aux oreilles de l'auditeur, le signal reçu

par chaque oreille ayant suivi un chemin différent. Si l'on observe une diminution de la corrélation interaurale, on perçoit alors un élargissement de la source.

3.4. BINAURAL

Grâce aux HRTF et par le biais d'algorithmes de calculs, on peut reproduire un signal multicanal sur deux canaux audio, au casque par exemple : on parle alors de binaural. Expérience déroutante lors des premières écoutes, mais si les HRTF choisies nous correspondent bien (la meilleure solution étant évidemment de les mesurer sur soi), cette technologie est puissante et promet un bel avenir.

Cette virtualisation au casque fait l'objet de nombreuses recherches actuellement, notamment au sein de grands organismes comme Radio France, France Télévisions, Orange Labs, Trinnov, etc.

N'ayant pas réalisé de tests binauraux en MPEG Surround, bien que ce codage soit compatible, je ne développerai pas davantage cette technique.

CHAPITRE 3 : LES FORMATS DE DIFFUSION AUDIO À LA TÉLÉVISION

1. DESCRIPTION DES PRINCIPAUX FORMATS AUDIO RENCONTRÉS À LA TÉLÉVISION

Il existe deux types de codages : les codages avec perte et les codages sans perte.

Avec les codages sans perte, appelés Lossless Coding en anglais, le décodeur fabrique un signal identique au signal source, caractère binaire pour caractère binaire. L'inconvénient majeur de ce type de codage est un faible gain de débit, car les taux de compression sont au maximum de l'ordre de 2 pour 1. Ce genre de codage est réversible, il est principalement utilisé en informatique.

Au contraire, dans un procédé de codage avec perte, appelé Lossy Coding en anglais, le décodeur ne délivre pas un signal identique à l'original. Ce type de procédé permet alors d'atteindre des taux de compression bien plus importants que ceux atteints par les codages sans perte, et sont donc fréquemment utilisés en audiovisuel.

Un codeur perceptuel analyse d'abord le signal original non compressé et le divise en sous-bandes (souvent en trente-deux sous-bandes). Chaque sous-bande est analysée avec un modèle perceptuel propre à chaque codec, qui permet alors d'adapter la quantification en fonction des informations utiles à coder. Une dernière étape facultative consiste en un codage entropique, il s'agit d'un codage informatique sans perte, qui exploite les redondances dans les données et les remplace par des codes plus courts, pour économiser encore du débit. Le codage de Huffman est souvent utilisé.

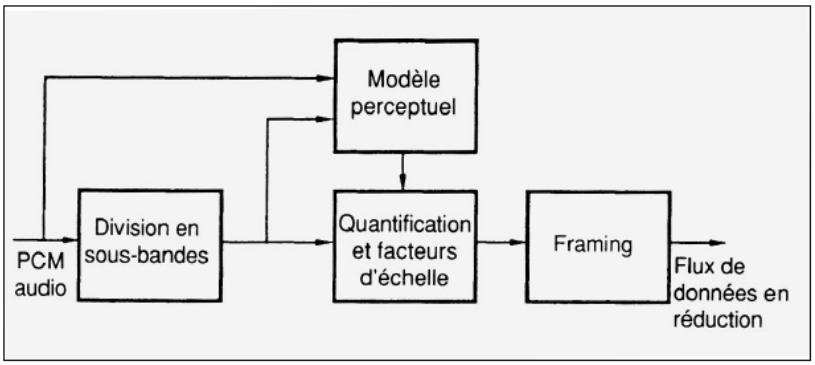


Figure 20 : Schéma de principe d'un codeur perceptuel

Les codeurs perceptuels, codeurs avec pertes, utilisent les caractéristiques de l'audition humaine, et exploitent les faiblesses de l'audition humaine, en ne codant que les informations qui seront perçues. Ces codeurs se basent sur les recherches en psycho-acoustique, et notamment sur les courbes isosoniques, sur le seuil absolu d'audition, et sur les masquages fréquentiels ou temporels, etc., pour proposer des réductions de débit plus ou moins importantes, tout en essayant de conserver la meilleure qualité possible.

1.1. MPEG-1

Le groupe MPEG, Moving Picture Expert Group, est un groupe de recherches sur le codage audio, né à la fin des années 1980, composé d'experts en images animées, issus d'organismes de normalisation, de laboratoires de recherches publiques et privées, ainsi que d'industriels.

Les normes MPEG audio définissent trois couches de codage, qui diffèrent par leur taux de compression. La norme de télévision numérique DVB-T préconise d'utiliser les couches I et II de la spécification MPEG-1 pour la transmission des signaux. Il existe quatre modes principaux que l'on peut choisir :

♪ Un mode « Stéréo » : les canaux gauche et droit sont codés de façon totalement indépendante.

♪ Un mode « Stéréo jointe » : les canaux gauche et droit sont comparés, la redondance entre ces deux voies est alors exploitée, afin de réduire le débit.

♪ Un mode « double canal» : les deux canaux audio sont indépendants (deux codages sont possibles : stéréo d'intensité ou MS).

♪ Un mode « mono » ou « simple canal » qui ne code qu'une seule voie.

Un codage en stéréo d'intensité utilise deux principes : pour les fréquences inférieures à 1,5 kHz, la stéréophonie est basée sur une différence de phase, tandis que pour les fréquences supérieures à 1,5 kHz, la stéréophonie est vraiment basée sur une différence d'intensité. Pour les très hautes fréquences, seuls les échantillons de sous-bande d'un seul canal sont transmis, avec des facteurs d'échelle différents en fonction du canal gauche ou droit.

Une trame MPEG Audio comporte un en-tête de trente-deux bits (header), qui contient la synchronisation et les informations système (format et couche, débit, fréquence d'échantillonnage, le mode de codage, et éventuellement un copyright) ; un code CRC (cyclic redundancy check ou contrôle de redondance cyclique) ou parité codé sur 16 bits pour détecter les erreurs, les données audio (de longueur variable selon la couche) et des données auxiliaires (cf. figure 21).

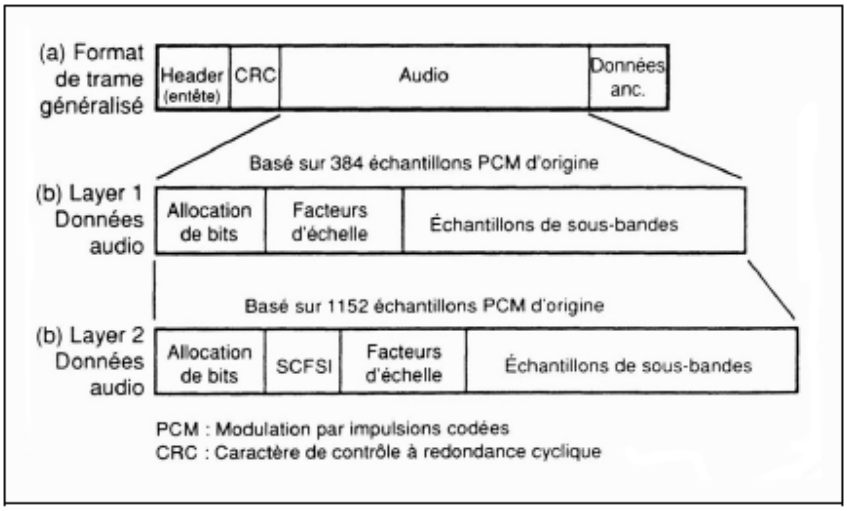


Figure 21 : Représentation d'une trame MPEG-1

La couche I ou pré-MUSICAM utilise l'algorithme PASC (Precision Adaptative Subband Coding) développé par Philips. Le débit est fixe, compris entre 32 et 448 kbits/seconde, une qualité haute fidélité nécessite un débit minimum de 192 kbits/seconde par canal audio (384 kbits/seconde en stéréophonie).

Le signal original est analysé par une banque de trente-deux filtres, ces trente-deux sous-bandes sont alors quantifiées, puis encodées en un flux MPEG audio. La quantification des coefficients de sous-bande est définie pour toute la trame, par un nombre codé sur quatre bits, chaque sous-bande est codée sur zéro à quinze bits, et le facteur d'échelle est codé sur six bits. Seul le seuil de masquage fréquentiel est utilisé.

Le champ d'allocation de bits/ESB, composé de trente-deux entiers codés sur quatre bits, définit la résolution de codage des échantillons de chaque sous-bande, comprise entre 0 et 15. Le champ «facteur d'échelle », composé de trente-deux entiers codés sur six bits, indique pour chaque sous-bande, le facteur multiplicatif des échantillons ainsi quantifiés.

Chaque trame contient 384 échantillons, soit douze échantillons par bande.

A 32 kHz, la trame dure 12 ms, à 44,1 kHz elle dure 8,7 ms et à 48kHz elle dure 8 ms.

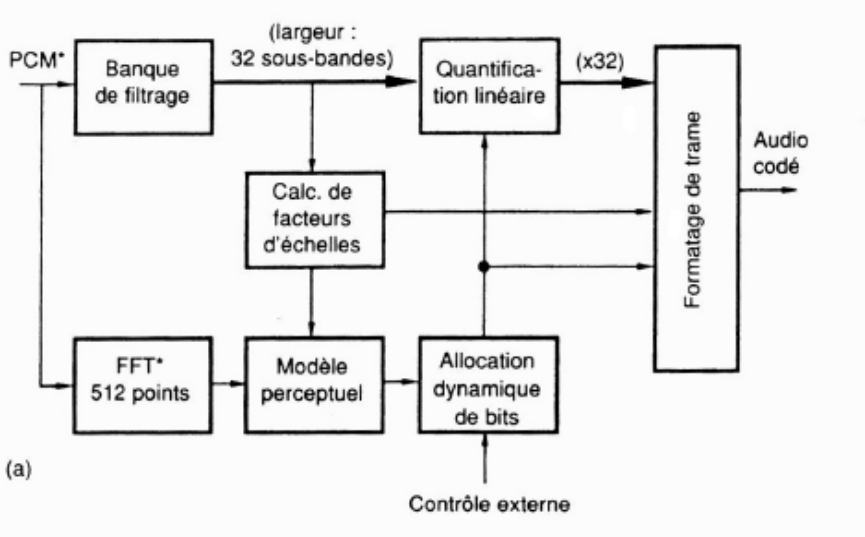


Figure 22 : Schéma de prinicpe d'un codeur MPEG-1 Layer-1

La couche II ou MUSICAM est le standard utilisé pour la radio (DAB) et la télévision (DVB) numériques terrestres européennes. La qualité est équivalente à celle de la couche I, mais avec une réduction de débit de 30 à 50%. Le débit est fixe, il peut être choisi entre 32 et 192 kbits/seconde par voie ; une qualité audio haute fidélité nécessite un débit minimum de 128 kbits/seconde, soit 256 kbits/seconde en stéréo. La trame a une durée triple que celle de couche I : il y a donc moins de bits « système », et la quantification des coefficients de sous-bande a une résolution dégressive (quantification sur quatre bits pour les bandes de basses fréquences, trois bits sur les bandes de moyennes fréquences et deux bits sur les bandes de haute fréquence). De plus, trois échantillons de sous-bande consécutifs peuvent être regroupés en « granules » pour être codés par un seul coefficient. La couche II exploite les seuils de masquage fréquentiel et temporel.

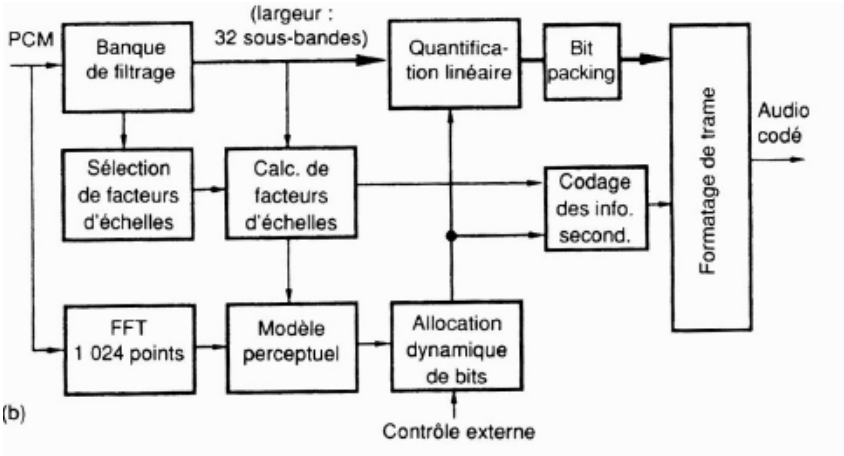


Figure 23 : Schéma de principe d'un codeur MPEG-1 Layer-2

La trame audio de cette couche II (cf. figure 21 page 51) contient un champ d'allocation de bits/ESB, constitué de trente-deux entiers codés sur deux à quatre bits selon la sous-bande, définit la résolution de codage des échantillons de chaque sous-bande et si ceux-ci sont regroupés par trois ou non. Le champ « SCFSI » (SCale Factor Selection Information) contient trente-deux entiers codés sur deux bits, qui indiquent si le facteur d'échelle de sous-bande s'applique à toute la trame ou s'il y a deux ou trois facteurs d'échelle. Le champ « facteur d'échelle » indique des nombres entiers codés sur six bits, qui sont les facteurs multiplicatifs des échantillons quantifiés pour la portion de

trame définie par le champ SCFSI. Chaque trame contient trois portions de douze échantillons pour chaque sous-bande, soit 1152 échantillons. La durée d'une trame de la couche II est trois fois plus longue que celle d'une trame de la couche I : 36 ms à 32 kHz, 26,1 ms à 44,1 kHz et 24 ms à 48 kHz.

Le format MPEG-1 Layer-2 est beaucoup utilisé dans la diffusion audio des chaînes de télévision, tous les programmes ont une intensité sonore égale à -23 LUFS, et le niveau n'est pas modifié jusqu'à réception du signal chez le téléspectateur.

La couche III définit le codage MP3, qui n'est pas utilisé à la télévision. Je ne le développerai donc pas.

1.2. MPEG-2

La norme MPEG-2 reprend le principe de la norme MPEG-1, en ajoutant une possibilité d'encodage d'un signal multicanal 5.0, tout en maintenant une rétro-compatibilité grâce à la création d'un downmix stéréophonique. De plus, des données auxiliaires spécifiques au MPEG-2 sont ajoutées, et sont ignorées par un décodeur MPEG-1. D'autres améliorations, comme un contrôle de la plage dynamique et une possibilité d'utiliser des fréquences d'échantillonnage divisées par deux, ce qui octroie alors une réduction de débit supérieure.

La figure 24 illustre le principe d'un codec MPEG-2 : un signal 5.0 (voire même 5.1) est matricé, un downmix stéréophonique est créé tandis que les cinq canaux sont encodés en un flux MPEG-2. Lors du décodage, le signal 5.0 (ou 5.1 selon les cas) est reconstruit.

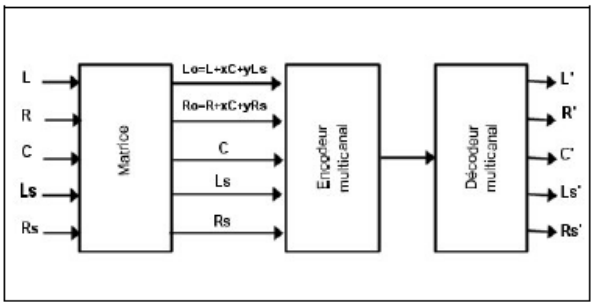


Figure 24 : Schéma de principe d'un codec MPEG-2

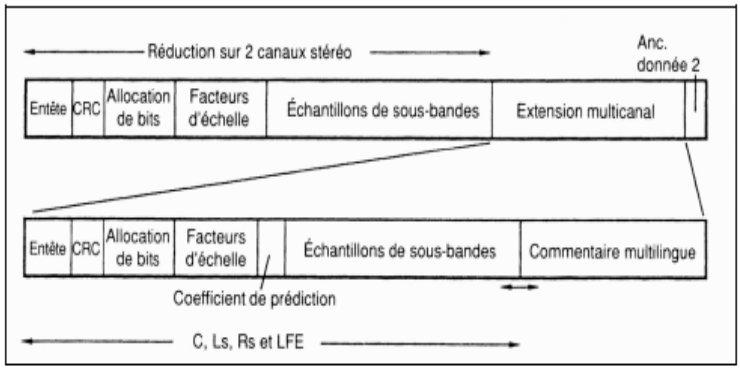


Figure 25 : Représentation d'une trame audio MPEG-2

Une trame MPEG-2 ressemble beaucoup à une trame MPEG-1, afin de maintenir une rétro-compatibilité, comme le montre la figure 25. La trame MPEG-2 contient un en-tête, un contrôle de redondance cyclique, un champ d'allocation de bits, un champ de facteurs d'échelle et les échantillons de sous-bandes qui correspondent au downmix stéréophonique réalisé lors du matriçage, ainsi que des données auxiliaires. Un champ d'extension multicanal est inséré entre les échantillons de sous-bandes du downmix et les données auxiliaires. Ce champ d'extension multicanal permet de coder les canaux centre, arrière gauche, arrière droit et LFE, avec un en-tête, un code de redondance cyclique, un champ d'allocation de bits, des facteurs d'échelle, un coefficient de prédiction, et les échantillons de sous-bandes, ainsi que des commentaires multilingues.

1.3. Dolby Digital et Dolby Digital Plus

1.3.1 Le Dolby Digital

Le format Dolby Digital est né fin 1991 et s'est démocratisé en 1995. C'est le premier format audio multicanal numérique, il est encore présent dans de nombreuses salles de cinéma et est utilisé pour les DVD. Le format Dolby Digital utilise l'algorithme de compression avec pertes « Audio Coding 3 » (AC-3). C'est un codage perceptuel basé sur les mêmes principes que le codage MPEG 2. Le Dolby Digital supporte différentes fréquences d'échantillonnage (32, 44,1 et 48 kHz) à différents débits (de 32 à 640 kbits/seconde), et il peut coder de l'audio monophonique, stéréophonique ou multicanal 5.1. L'audio codé en Dolby Digital sur DVD a un débit de 448 kbits/seconde, tandis qu'il peut monter jusqu'à 640 kbits/seconde pour un disque Blu-Ray.

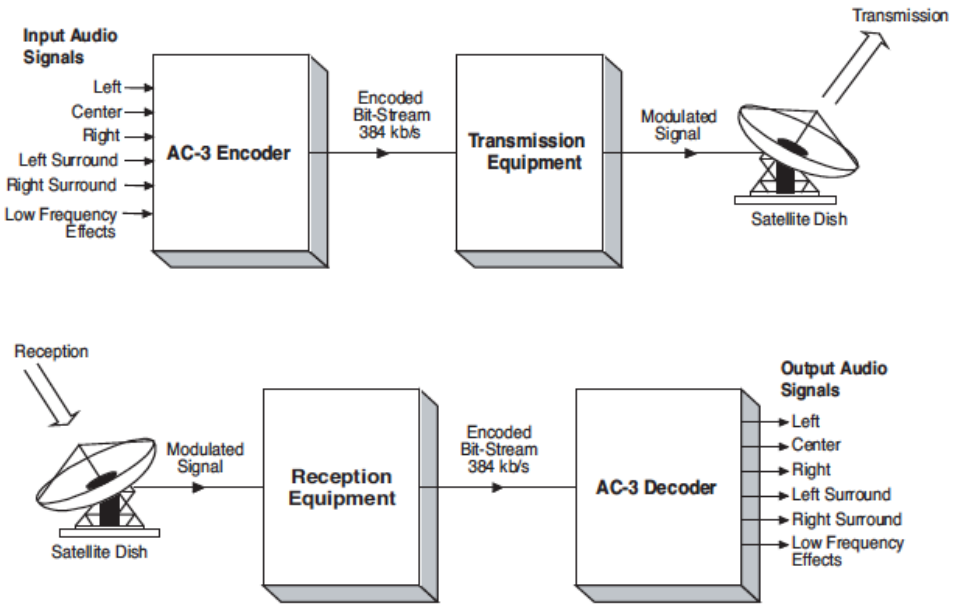


Figure 26 : Diffusion audio par satellite en AC-3

Le Dolby Digital encode un signal multicanal 5.1 en six canaux discrets : cinq canaux sonores ayant une bande passante de 20 Hz à 20 kHz, et un canal LFE réservé aux basses fréquences, filtré à 120 Hz. Les six canaux se répartissent ainsi : deux canaux avant frontaux, gauche et droite, principalement pour les ambiances, les effets et les musiques ; un canal central pour les dialogues ; deux canaux arrière gauche et droite pour des ambiances, des effets, voire de la musique et autres effets (telle qu'une réverbération) ; et un canal LFE, utilisé pour renforcer les basses de 3 à 120 Hz). Contrairement aux codages matricés, les six canaux sont indépendants.

Tableau 1 : Trame audio AC-3

Sync	CRC 1	Sy. info	BSI	Audio 1	Audio 2	Audio 3	Audio 4	Audio 5	Audio 6	Aux	CRC 2

Une trame AC3 comporte entre autre un champ de synchronisation, ainsi que deux champs CRC ou Cyclic redundancy check ou contrôle de redondance cyclique, qui permettent de détecter les erreurs dans une trame, l'un au début de la trame, l'autre à la fin de la trame. Le champ BSI contient toutes les données relatives à l'audio, telles que la

fréquence d'échantillonnage, le débit, le nombre de canaux encodés, la compression utilisée, le dialnorm, le lieu où ça a été mixé (Room Type), les services disponibles (DRC, Dynamic Range Control ou contrôle de dynamique, mode karaoké, dialogue, voice-over, etc.).

La métadonnée Dialnorm, ou Dialogue Normalization, ainsi nommée dans les paramètres, correspond au niveau moyen du dialogue, mesuré sur une certaine durée avec l'algorithme Leq(A). Ce nombre, positif, est généralement compris entre 1 et 31 et représente l'écart entre la modulation maximale (c'est à dire une modulation de 100% qui correspond à un niveau de 0 dBFS en numérique) et le niveau moyen du dialogue (appelé Dialog Level). Il est important de renseigner la métadonnée Dialnorm avec la véritable valeur, afin que le décodeur Dolby Digital, dont le niveau moyen de dialogue de référence est -31 dBFS, puisse appliquer la véritable atténuation nécessaire. Cette valeur de Dialnorm se rapproche de la valeur d'intensité sonore d'un programme (ou integrated loudness). Désormais, la valeur attendue dans le champ Dialnorm est donc la valeur « integrated loudness » mesurée sur l'intégralité du programme (soit -23 LUFS, à +/-1 LU normalement). Le décodeur Dolby se charge ensuite d'appliquer au signal une atténuation (environ égale à -8 dB), afin que le dialogue soit normalisé à sa valeur de référence, comme le montre la figure 27bis.

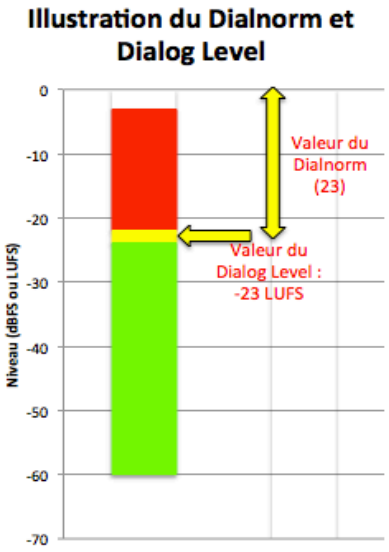


Figure 27 : Illustration du Dialnorm

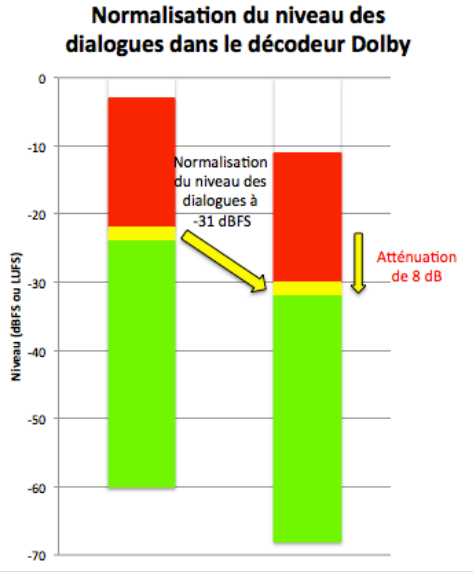


Figure 27 bis : Normalisation du niveau des dialogues

Avant l'encodage d'un programme, il faut déterminer les paramètres de compression. Il existe six profils de compression prédéfinis, illustrés sur la figure 28. Le profil nommé « Null band » définit une zone linéaire, où il n'y a pas de compression, centrée sur la valeur du Dialog level, et dont la largeur est définie par le technicien. Il existe un profil « Boost Range » qui sert à amplifier les bas niveaux, tandis que les deux profils « Early Cut Range » et « Cut Range » servent à compresser les forts niveaux. Un profil nommé « None » permet de ne sélectionner aucun profil.

Au décodage, selon le type de décodeur à six canaux qu'il possède, le téléspectateur peut choisir ou non d'activer le Dynamic Range Control, selon qu'il souhaite entendre un signal légèrement compressé (Line Mode, les dialogues sont reproduits à un niveau de -31 dBFS) ou un signal ultra compressé (RF Mode, les dialogues sont reproduits à -20 dBFS et la dynamique est très réduite). En revanche, s'il ne dispose que d'un décodeur deux canaux, le Dynamic Range Control est forcément activé, afin d'éviter des saturations. Avec les nouvelles réglementations de niveau, en mode RF, le niveau des dialogues est tout de même normalisé à -31 dBFS dans le décodeur Dolby, mais la plage de dynamique est beaucoup plus restreinte.

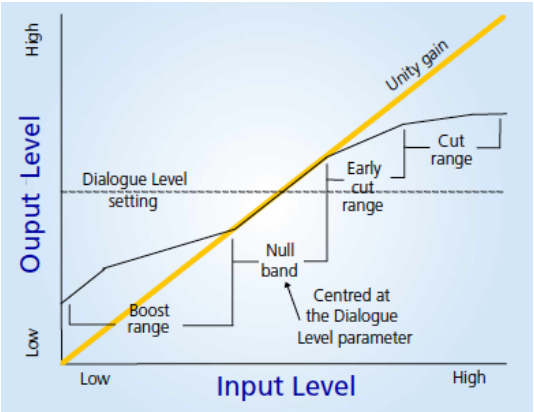


Figure 28 : Profils DRC du format Dolby Digital

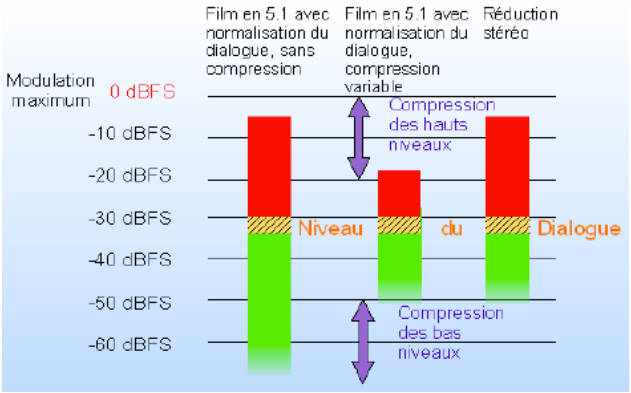


Figure 29 : Illustration du fonctionnement du DRC

D'autres métadonnées qui concernent le downmixing sont ajoutées. Il en existe plusieurs types : un downmixing LoRo (Left Only, Right Only), qui est le plus utilisé et consiste à sommer dans le canal Lo le canal gauche à 0 dB, le canal central à -3 dB et le canal arrière gauche à -3 dB, et dans le canal Ro le canal droit à 0 dB, le canal central à -3dB et le canal arrière droit à -3 dB[18] ; le downmixing LtRt (Left Total Right Total) est quant à lui compatible avec les décodeurs Dolby Pro Logic I et II et permet d'être écouté soit en stéréo, soit en quatre canaux gauche, centre, droit, surround si l'on dispose d'un décodeur Dolby Pro Logic, ou encore un downmixing Mono, les canaux Lo et Ro sont alors sommés. Les métadonnées contiennent alors tous ces paramètres, ainsi que le downmixing préconisé par le technicien. Les téléspectateurs équipés d'un décodeur Dolby Digital 2.0 pourront choisir le downmixing qu'ils souhaitent si le décodeur le propose, sinon c'est le downmixing préconisé par le technicien qui sera lu.

En Dolby Digital, les six canaux sont encodés séparément, chaque bloc audio contient 256 échantillons, soit un total de 1536 échantillons par trame. La durée effective de la trame dépend évidemment de la fréquence d'échantillonnage et du débit alloué. Un champ Aux contient des données auxiliaires.

Le Dolby Digital permet une compression allant jusqu'à un ratio de 15 :1, la compression est plus importante que celle proposée par le format DTS.

1.3.2. Le Dolby Digital Plus

Le Dolby Digital Plus, également appelé e-AC3 (Enhanced Audio Coding 3), est une évolution du codage Dolby Digital. Il propose des améliorations, telle qu'une meilleure qualité pour un débit équivalent, et autorise une transmission de treize canaux, et d'un canal LFE. Cependant, il n'est pas rétro-compatible. Le Dolby Digital Plus permet d'encoder des signaux 1.0 à 13.1, il fonctionne à différentes fréquences

[18] NB : les coefficients du downmixing sont donnés à titre indicatif. Le technicien peut en choisir d'autres.

d'échantillonnage (32 kHz, 44,1 kHz et 48 kHz), sa résolution peut atteindre 24 bits, et son débit est compris entre 32 kbits/seconde et 6144 Mbits/seconde.

Ce codage est particulièrement efficace puisqu'il permet de transporter la version française 5.1 multicanal en six canaux discrets avec une meilleure qualité, tout en intégrant, comme son prédécesseur, les métadonnées Dialog Level, Dynamic Range Control (DRC) et les paramètres de downmixing pour les téléspectateurs qui ne possèdent qu'un décodeur Dolby Digital Plus 2.0 et/ou pour ceux équipés d'un seul système stéréophonique.

Les premières chaînes diffusées en haute définition sur la télévision numérique terrestre émettaient un signal audio 5.1 en Dolby Digital. En juin 2009, les chaînes TF1 HD, France 2 HD et M6 HD ont changé de format audio, évoluant vers le Dolby Digital Plus afin d'améliorer la qualité et d'optimiser la bande passante. Ce changement a alors nécessité une mise à jour des décodeurs compatibles, voire un changement de décodeur pour certains foyers dont le décodeur était incompatible en e-AC3. Le Dolby Digital Plus prend en compte beaucoup de paramètres, aujourd'hui essentiels pour une diffusion télévisuelle de qualité, c'est pourquoi ce format est majoritairement utilisé par les chaînes en haute définition pour diffuser un signal 5.1.

1.4. MPEG-4 : ADVANCED AUDIO CODING

Le codage Advanced Audio Coding a été standardisé en 1997 par la société Fraunhofer IIS, en tant qu'extension de la norme MPEG-2, pour concurrencer le MP3. Ce format est plus performant pour coder les signaux stationnaires avec des blocs de 1024 échantillons, tandis que les transitoires sont codées avec des blocs de 128 échantillons. Par contre, ce format n'est pas rétro-compatible. Le format MPEG-2 AAC supporte jusqu'à 48 canaux, des fréquences d'échantillonnage comprises entre 8 et 96 kHz (contre 16 à 48 kHz pour le mp3), les fréquences au-delà de 16 kHz sont mieux gérées, le mode

« Stéréo jointe » est plus souple, et permet une gestion des droits numériques (DRM), pour contrôler l'utilisation des fichiers dans ce format.

La version standard comporte trois profils, ayant des niveaux de complexité différents : Main Profile (le plus complet), Low Complexity Profile (LC, le plus utilisé), et Scalable Sampling Rate Profile (SSR, qui autorise un codage hiérarchique, c'est-à-dire qu'il autorise la transmission dans un même flux de tous les éléments correspondants aux différents niveaux de qualité).

Puis en 1999, la première version de la norme MPEG-4 est créée. La version AAC de la norme MPEG-2 est alors améliorée pour devenir un codage audio de haute qualité sous la norme MPEG-4. De nouveaux outils sont intégrés, telle que la PNS (Perceptual Noise Substitution ou substitution du bruit perçu : puisque l'on a constaté qu'un « bruit » est similaire à un autre, et donc que sa structure véritable a peu d'importance, la technique de la PNS consiste à remplacer les éléments identifiés comme du « bruit » par un signal de bruit généré par le décodeur, ce qui offre un réel gain de débit au codage). Un nouvel outil de prédiction, appelé long term prediction ou LPT, requiert peu de ressources processeurs, tout en étant aussi efficace que l'outil de prédiction du format MPEG-2 AAC.

Il faut par contre se méfier : un codeur perceptuel fonctionne très bien dans une certaine plage de débits, mais si on veut descendre en dessous d'un certain seuil, de nombreux artefacts apparaissent, comme notamment une réduction de la bande passante ou une réduction de l'image stéréophonique.

1.5. MPEG-4 : High Efficiency Advanced Audio Coding

Le format MPEG-4 AAC a connu une évolution majeure en 2003, avec l'ajout de la technologie SBR (Spectral Band Replication) pour créer un nouveau format nommé HE-AAC (High Efficiency Advanced Audio Coding ou AAC Plus).

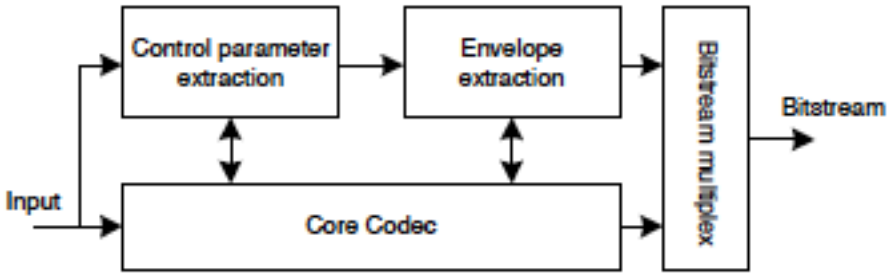


Figure 30 : Schéma de principe d'un encodeur SBR

La technologie SBR ou reconstruction de bande spectrale repose sur la forte corrélation qui existe dans un signal audio entre les basses fréquences et les hautes fréquences. Il est donc alors inutile de coder les hautes fréquences, mais seulement de transmettre des données complémentaires à faible débit, qui décrivent ces fréquences, pour permettre la meilleure reconstruction possible. En revanche, les basses fréquences sont codées par le codeur AAC standard. Le flux final issu de l'encodeur contient les données issues du codeur AAC pour le bas du spectre, et les données auxiliaires SBR.

Sur la figure 33 (page 63), on observe une reconstruction du signal sans ajustement de l'enveloppe : le signal reconstruit ne correspond pas au signal original, les hautes fréquences n'ont pas la bonne amplitude. Un décodeur SBR a donc besoin des données auxiliaires qui contiennent des informations sur l'enveloppe des hautes fréquences notamment, pour pouvoir reconstruire un signal le plus semblable possible à l'original (cf. figure 34 page 63).

De plus, pour les signaux dont les basses fréquences sont peu corrélées aux hautes fréquences, les données auxiliaires en tiennent compte, et comportent davantage d'informations concernant les hautes fréquences.

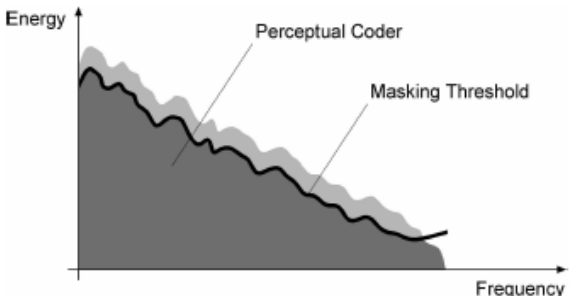


Figure 32 : Représentation d'un signal et du seuil de masquage

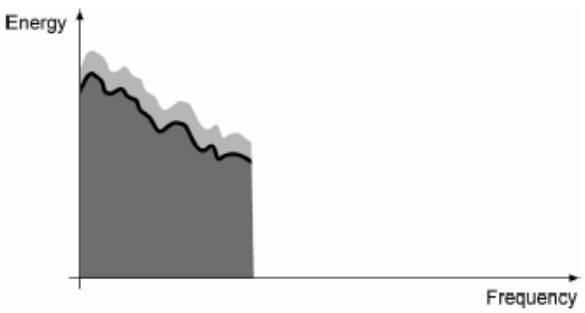


Figure 31 : Filtrage des hautes fréquences par la technologie SBR

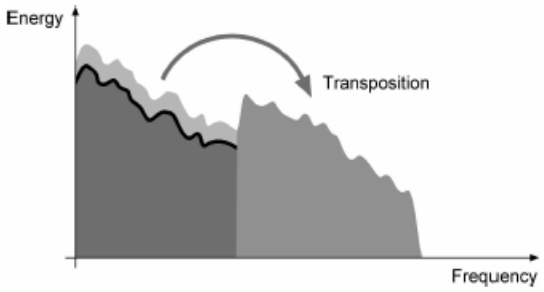


Figure 33 : Reconstruction des hautes fréquences basée sur le seul contenu du bas du spectre

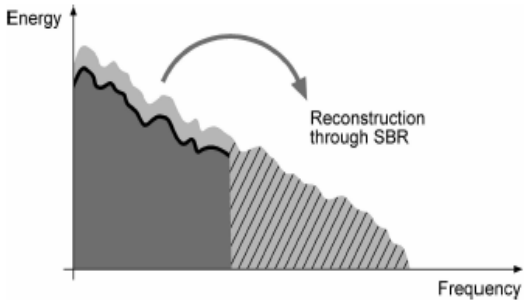


Figure 34 : Reconstruction du haut du spectre par la technologie SBR (basée sur le contenu du bas du spectre + ajustement de l'enveloppe)

La version MPEG-4 HE-AAC a subi une autre évolution en 2004, en ajoutant la technologie Parametric Stereo (PS ou Stéréo Paramétrique). Cet ajout a donné naissance à un nouveau codec appelé MPEG-4 HE-AAC v2 (ou AACPlus v2). Le codage Stéréo Paramétrique consiste à extraire les informations spatiales du signal stéréophonique, de coder le signal stéréophonique en un signal monophonique, et d'encoder les informations spatiales dans un flux de données auxiliaires, à bas débit.

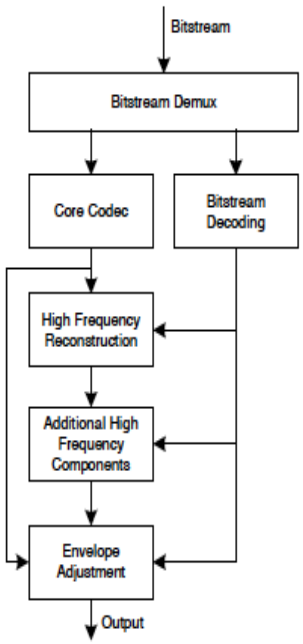


Figure 35 : Schéma de principe d'un décodeur SBR

Les informations paramétriques se basent sur trois indices :

- ♪ la différence interaurale de niveau (ILD).
- ♪ la différence interaurale de phase (ITD).
- ♪ la cohérence entre les deux signaux (ICC)[19].

[19] Cf. Chapitre 2, 3.1 Page 45.

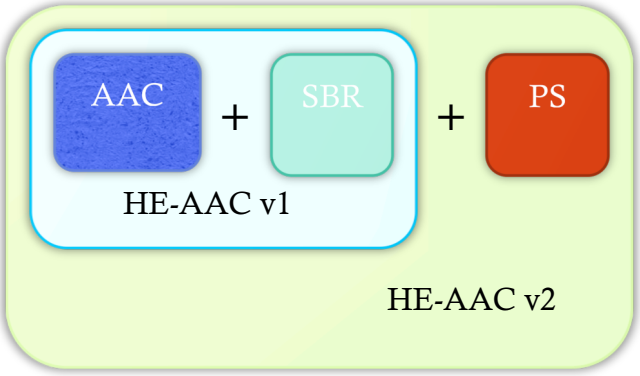


Figure 36 : Améliorations de la norme MPEG-4

Les formats AAC, HE-AAC v1 et HE-AAC v2 supportent des métadonnées, comme le Dolby Digital Plus, seuls les noms changent, comme le montre le tableau suivant.

Tableau 2 : Comparaison des métadonnées des codecs MPEG-4 et AC-3

Paramètres	AAC et HE-AAC	AC-3 et E-AC-3
Normalisation loudness	Program Reference Level	Dialnorm
Dynamic Range Control Faible compression Forte compression	Dynamic Range Control Compression value	Line Mode RF Mode
Downmix	Matrix-mixdown Downmixing levels	Downmix

De plus, on peut aussi introduire les métadonnées relatives aux mesures loudness du programme. Comme en Dolby Digital Plus, toutes les métadonnées devraient provenir de la production, être incluses dans le flux Dolby E, puis être transcodées pour correspondre aux métadonnées attendues par un codec MPEG-4 et ensuite être transmises à l'encodeur HE-AAC ou AAC.

Un récepteur compatible HE-AAC et capable de décrypter les métadonnées (comme par exemple les set-top boxes, cf. figure 38) décodent le flux HE-AAC en

appliquant les paramètres inclus dans les métadonnées, en un signal audio PCM. Le décodeur peut même créer un downmix stéréophonique si le système audio n'est pas compatible en 5.1. En revanche, les décodeurs qui ne sont pas compatibles avec le flux de métadonnées liront le signal HE-AAC tel quel.

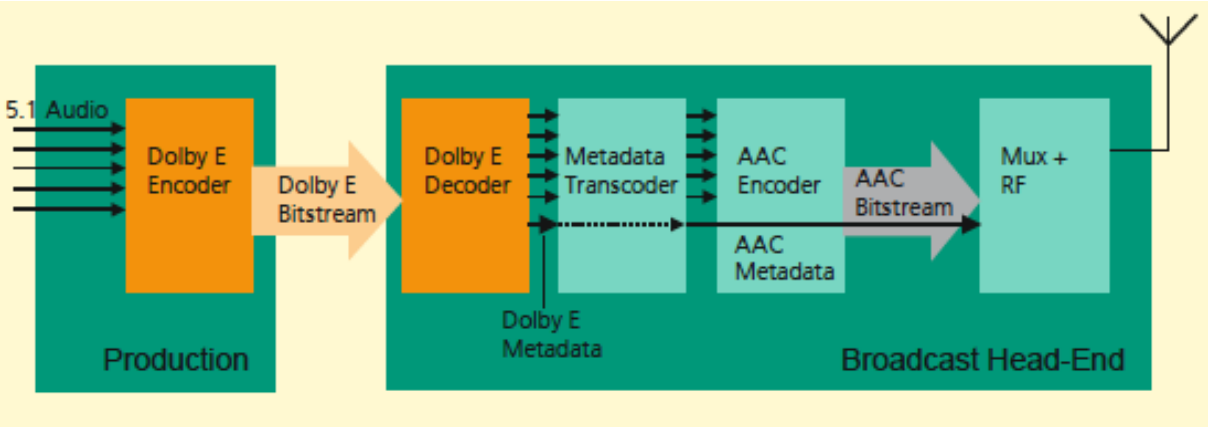


Figure 37 : Chaîne de diffusion, avec transcodage des métadonnées incluses dans le Dolby E pour créer un flux MPEG-4

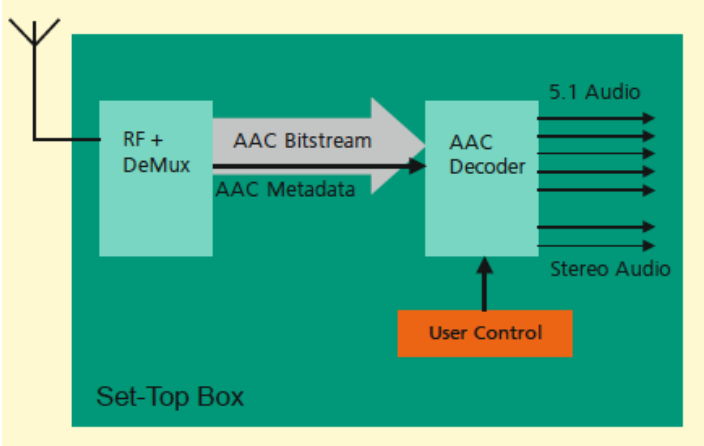


Figure 38 : Récepteur compatible HE-AAC

2. Quels Formats Audio Diffusés pour Quelles Chaînes et sur Quel Vecteur de Diffusion ?

2.1. Télévision Numérique Terrestre en Définition Standard (TNT SD)

La régie finale de chaque chaîne assemble les programmes (image, son, habillage, sous-titrage, etc.), puis fournit les signaux audio et vidéo non compressés (en PCM linéaire pour le son) via une liaison SDI[20] à un débit de 270 Mbits/seconde, à la tête de réseau nationale, gérée par TDF[21]. Cette société se charge alors du codage source des programmes et du multiplexage des chaînes, avant de les diffuser. Le codage permet de garantir une qualité homogène des programmes au sein d'une chaîne et d'économiser du débit.

En télévision numérique terrestre, pour les chaînes gratuites en définition standard, l'image est principalement diffusée en format MPEG-2, tandis que l'audio est souvent un flux stéréophonique, encodé en canaux discrets en MPEG-1 Layer-2 à 192 kbits/seconde.

Pour les chaînes payantes diffusées en définition standard, l'image est diffusée en MPEG-4, tandis que le format audio est souvent du MPEG-1 Layer-2 ou de l'HE-AAC.

Quelque soit le format, et ce depuis le 1er janvier 2013, tous les programmes doivent avoir une intensité sonore de -23 LUFS à +/- 1 LU en sortie de régie, le niveau n'est pas modifié lors de l'encodage en MPEG-1 Layer-2.

[20] SDI : Serial Digital Interface ou Interface Numérique Série, c'est un protocole de diffusion ou de transport des différents formats de vidéo numérique, avec possibilité d'intégrer l'audio. Chaque image comporte 625 lignes, mais seulement 576 lignes utiles pour la vidéo, il y a alors 25 lignes supprimées pour une trame et 24 lignes pour l'autre trame. Ces suppressions de lignes permettent de contenir jusqu'à 16 canaux audio encodés en 48 kHz – 24 bits.

[21] TDF : TéléDiffusion de France.

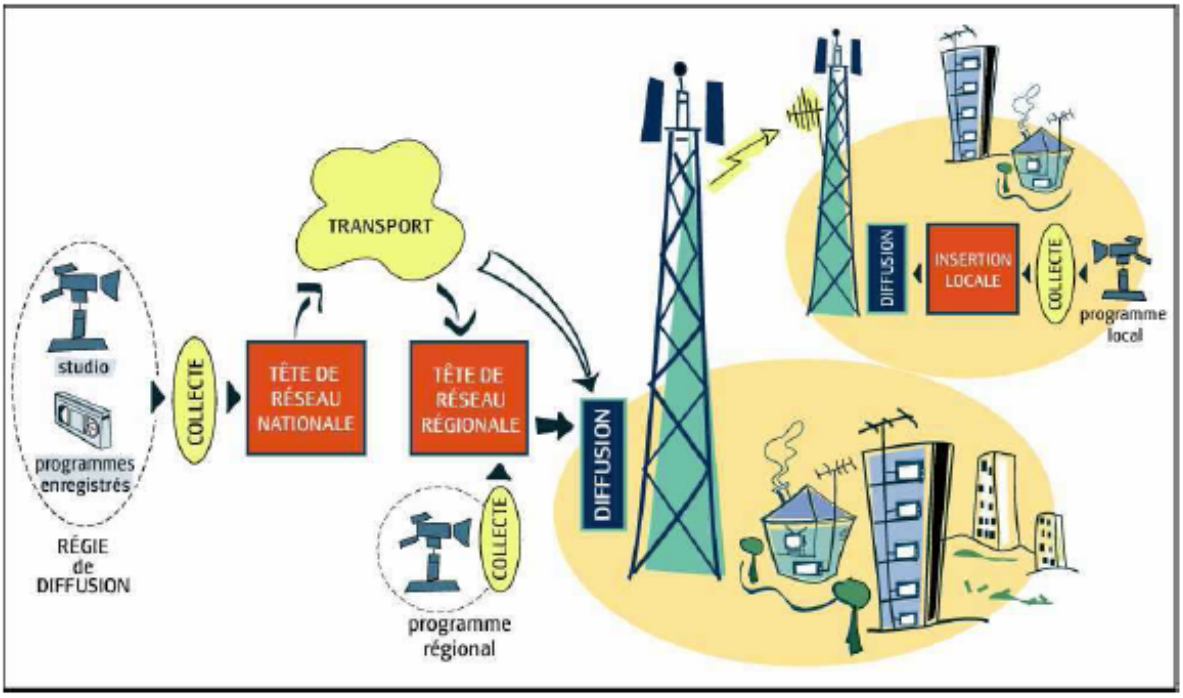


Figure 39 : Schéma de diffusion de la TNT SD, de la régie finale au téléspectateur

Si le téléspectateur écoute la télévision sur les haut-parleurs du téléviseur, ou sur deux enceintes externes, le niveau n'est pas modifié. En revanche, si le téléspectateur dispose d'un amplificateur home-cinéma avec décodeur intégré, le décodeur doit être appliquer une atténuation de 8 dB aux signaux MPEG-1 Layer-2 des chaînes en définition standard, afin d'homogénéiser les niveaux entre les flux Dolby Digital Plus des chaînes en haute définition, dont les dialogues sont normalisés à -31 dBFS et les niveaux des flux MPEG-1 Layer-2 des chaînes en définition standard, égaux normalement à -23 LUFS. Tous les nouveaux équipements depuis 2012 doivent appliquer cette atténuation, les anciens décodeurs ne l'appliquent pas, sauf si une mise à jour est possible.

♪ Exemple 1 : France 2

Depuis Octobre 2012, en sortie de la régie finale de France 2, le signal disponible en définition standard est un signal SDI (Y, Cr, Cb) à 625 lignes/50i, avec un débit de 270 Mbits/seconde, qui contient deux groupes audio déclarés SMPTE-272M. Chaque groupe audio contient deux paires AES, où le signal est échantillonné à 48 kHz, avec une

résolution de 20 bits, l'audio est « embedded » ou intégré sur les données auxiliaires de toutes les lignes.

La première paire AES contient le canal audio en PCM en stéréo jointe de la version française du programme. La deuxième paire AES transporte le son en PCM pour l'audio-description. La troisième paire AES comporte le son en PCM en version multilingue (c'est-à-dire la version original dans la majorité des cas). Si le programme ne contient ni audio-description, ni version multilingue, les deuxième et troisième paires AES contiennent alors le son stéréophonique PCM en version française.

Le son est alors diffusé après traitement audio Jünger (processeur d'antenne) et tatouage numérique (watermarking). La société TDF ré-encode tous les flux stéréophoniques en MPEG-1 Layer-2 à 192 kbits/seconde pour la diffusion des programmes de France 2 en définition standard.

Les sous-titrages sont produits selon la norme européenne WST (World System Teletex) ou plus communément nommée Télétexte européen, avec l'aide du système Ceefax. D'autres chaînes, comme TF1, France 3, France 5, Arte, M6 et Canal+, proposent des sous-titrages, avec le même protocole. Les sous-titrages sont stockés en lignes 10 et 323. On retrouve le sous-titrage pour sourds et malentendants en page 888, tandis que les sous-titrages en français (pour la version multilingue) se situent en page 889.

♪ Exemple 2 :

Les chaînes France 3, France 4, France 5 et France Ô ne diffusent qu'en définition standard.

Pour France 3 et France 5, la première paire AES comporte un signal stéréophonique de la version française, tandis que la deuxième paire AES contient le son PCM stéréophonique pour l'audio-description. Les troisième et quatrième paires AES ne contiennent pas d'audio.

Pour France 4, la première paire AES contient le signal stéréophonique de la version française, et la deuxième paire AES contient l'audio en PCM de la version

originale (souvent en langue anglaise). Les troisième et quatrième paires AES ne contiennent pas d'audio.

Enfin, en ce qui concerne France Ô, seule la première paire AES est utilisée : elle contient d'audio en PCM de la version française.

Tous les flux audio stéréophoniques de ces chaînes sont encodés en MPEG-1 Layer-2 à 192 kbits/seconde.

2.2. Télévision Numérique Terrestre en Haute Définition (TNT HD)

De même, les régies finales des chaînes en haute définition assemblent leurs programmes (image, son, habillage, etc.) et transmettent les signaux audio et vidéo via une liaison HD-SDI[22] à un débit de 1,485 Mbits/seconde à la tête nationale de réseau, gérée par TDF. La tête nationale réseau compresse la vidéo en MPEG-4 AVC HD, en moyenne à 7 Mbits/seconde et transcode l'audio issu du Dolby E intégré dans le flux HD-SDI en Dolby Digital Plus.

L'image des chaînes en haute définition est souvent encodée en format MPEG-4 (H.264), tandis qu'elles sont tenues de diffuser trois flux audio différents : un flux en version française, un flux en version originale (lorsqu'elle existe) et un flux d'audio-description (lorsque les programmes ont été audio-décrits). L'audio diffusé pour la version française est généralement un signal multicanal 5.1, encodé en six canaux discrets en Dolby Digital Plus à 256 kbits/seconde, mais peut aussi être un signal monophonique ou stéréophonique (selon le programme).

[22] HD- SDI : High Definition Serial Digital Interface ou Interface Numérique Série Haute Définition, c'est un protocole de transmission semblable au protocole SDI, mais appliqué à la transmission de signaux vidéo haute définition. Les images transmises peuvent être en 720p ou 1080i, le débit peut aller jusqu'à 1,485 Gbits/seconde. Huit canaux audio stéréophoniques peuvent être transmis en même temps, échantillonnés à 48kHz en 24 bits.

Les signaux en Dolby Digital Plus intègrent plusieurs métadonnées : le Dialnorm, le Dynamic Range Control et les paramètres de downmixing. Désormais, tous les programmes télévisés doivent avoir une intensité sonore (ou « integrated loudness ») égale à -23 LUFS à +/-1 LU, et c'est cette valeur d'intensité sonore mesurée que l'on doit renseigner dans le champ Dialnorm. Le décodeur Dolby se charge ensuite de lui appliquer l'atténuation nécessaire pour arriver à son niveau de référence, fixé à -31 dBFS.

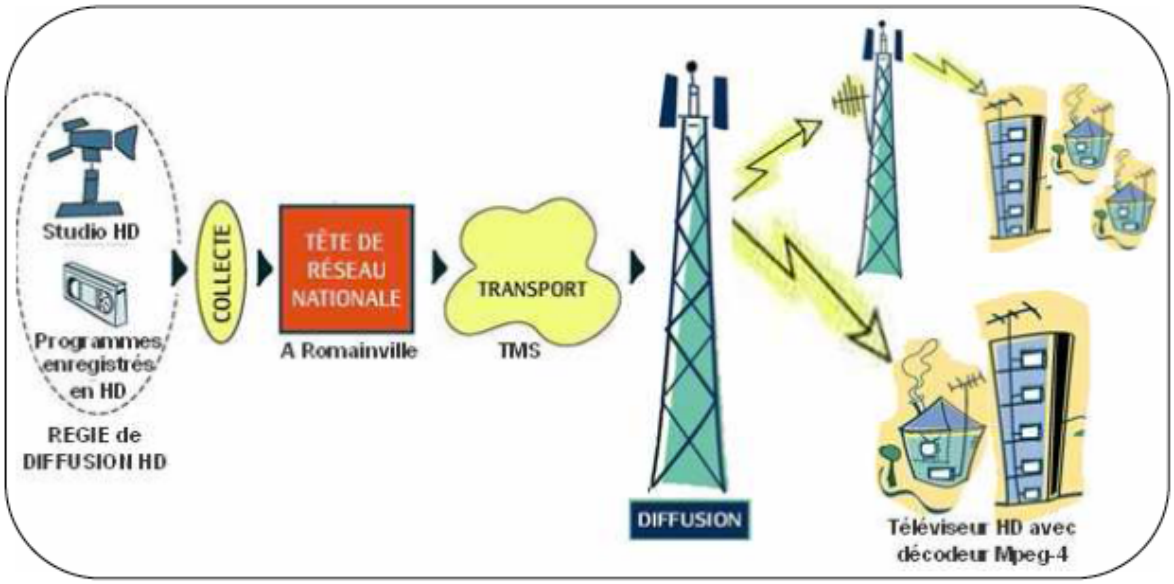


Figure 40 : Schéma de principe de la TNT HD, de la régie finale au téléspectateur

Pour les téléspectateurs qui écoutent le flux Dolby Digital Plus 2.0 sur les haut-parleurs du téléviseur ou sur deux enceintes externes, ils utilisent la sortie analogique du décodeur, en line mode. Le décodeur doit alors appliquer un gain positif de 8 dB au flux e-AC3, pour qu'il soit au même niveau, c'est-à-dire à -23 LUFS, que les flux MPEG-1 Layer 2 des chaînes diffusées en définition standard. Les nouveaux décodeurs appliquent ces préconisations, néanmoins les anciens décodeurs qui ne supportent pas de mise à jour, n'appliquent aucun gain.

Les flux audio multicanaux peuvent être émis en Dolby Digital Plus ou en HE-AAC, les récepteurs doivent alors être aptes à réaliser un downmix stéréophonique. Le

format HE-AAC permet lui aussi d'intégrer des métadonnées, semblables à celles du Dolby Digital.

♪ Exemple : France 2 HD

Depuis Octobre 2012, la régie finale de France 2 délivre, à destination d'une diffusion haute définition, un signal HD-SDI (Y, Cr, Cb) ayant une résolution de 1920*1080 pixels/50/2 :1, avec un débit total de 1,485 Gbits/seconde. L'audio est « embedded » ou intégrée sur les données auxiliaires des lignes Cr et Cb. Deux groupes audio, de deux paires AES chacun, sont déclarés SMPTE 299M, l'audio est échantillonné à 48 kHz, en 24 bits pour chaque paire AES.

La première paire AES contient le son stéréophonique en PCM en version française. La deuxième paire AES comporte le Dolby E. La troisième paire AES transporte le son PCM pour l'audio-description, et la quatrième paire contient le son en PCM de la version multilingue (version originale dans la plupart des cas). Si le programme ne dispose pas d'audio-description ni de version multilingue, les paires 3 et 4 contiennent alors le son PCM en version française.

Juste avant de sortir de la régie finale, le son traverse un processeur d'antenne de type Jünger, un up-mixeur et subit un tatouage numérique. En effet, la régie finale de France 2 n'était pas adaptée pour transporter et diffuser un son 5.1 au moment où la chaîne a commencé la diffusion en Haute Définition. C'est donc la version stéréophonique qui est transportée, et qui est up-mixée.

TDF encode le flux audio issu du dolby E en signal 5.1 en Dolby Digital Plus à 256 kbits/seconde ; les flux stéréophoniques de la version multilingue, de l'audio-description (et éventuellement la version française 2.0 de secours) sont encodés en Dolby Digital Plus en 2.0. Le nouveau centre d'échange et de diffusion, qui regroupera les régies finales de France 2, France 3, sera effectif fin 2013, et dès lors, un véritable signal multicanal 5.1 sera diffusé en Haute Définition.

Le sous-titrage OP 47, pour sourds et malentendants, et le sous-titrage en français pour la version multilingue sont intégrés aux lignes 10 et 571.

2.3. Fournisseurs d'Accès à Internet, Câble et Satellite

La régie finale des chaînes fournit aussi le flux direct en liaison SDI au satellite (via France Télécom et CanalSat), aux opérateurs du câble et aux fournisseurs d'accès à Internet, via une liaison SDI à 270 Mbits/seconde. Ces opérateurs sont ensuite libres de les ré-encoder dans le format qu'ils souhaitent, pour les diffuser en direct, avec quelques secondes de décalage par rapport à la TNT, d'où de grandes disparités de qualité et de niveaux sonores observées selon le vecteur de diffusion regardé.

La diffusion de la chaîne France 3 est un cas particulier avec les décrochages régionaux, qui concernent principalement les journaux télévisés, certaines émissions locales le week-end et certains évènements sportifs. La chaîne nationale est diffusée depuis la régie finale à Paris sur les canaux satellite, du câble et des fournisseurs d'accès à Internet. Les vingt-quatre centres régionaux de France 3 récupèrent le signal national par satellite, et se chargent alors de l'assemblage du contenu national et des contenus locaux, et l'injectent dans les émetteurs TNT de la région, gérés par l'opérateur TDF. Les flux télévisés des centres régionaux sont alors remontés via des lignes spécialisées sous protocole Internet à Paris, et alimentent à leur tour les fournisseurs d'accès à internet, ainsi que l'opérateur câblé Numericable, qui proposent toutes les déclinaisons régionales de France 3 dans leur bouquet, en plus de la chaîne nationale. Ces signaux télévisés sont alors réinjectés dans un satellite en haute qualité, pour être ensuite récupérés par les quarante-neuf antennes régionales, et émises sur la TNT locale.

2.4. TÉLÉVISION CONNECTÉE, TÉLÉVISION DE RATTRAPAGE

Les services de télévision connectée, de télévision de rattrapage et de vidéo à la demande sont généralement réalisés par des filiales ou des prestataires de service, à la demande des chaînes.

Chez France Télévisions, il existe un portail de visionnage en direct sur internet et de télévision de rattrapage pour toutes les chaînes du groupe (France 2, France 3, France 4, France 5, France Ô et la 1ère) : le portail nommé Pluzz, existe depuis juillet 2010. La majorité des programmes diffusés, à savoir téléfilms, fictions, documentaires, magazines, journaux télévisés, jeux, divertissements, sont disponibles gratuitement durant sept jours après la diffusion. En Février 2012, on dénombrait 1200 programmes en stock par jour, ce qui correspond à un buffer d'antenne de sept jours.

Il existe trois versions de ce portail : une version web, une version smartphone/tablette (via l'application Francetv Pluzz) et une version pour les boxes des fournisseurs d'accès à Internet.

Derrière ce portail se cache une plateforme de gestion et de diffusion, qui se charge dans un premier temps de récupérer les flux : enregistrement des flux pour la diffusion en direct, récupération des programmes via un serveur et numérisation des programmes livrés en betacam. Il y a aussi un système de gestion des droits de diffusion, chaque programme étant associé à des droits accordés par les éditeurs : en fonction des supports (téléviseur, mobile, web), des vecteurs de diffusion (TNT, FAI, satellite), de la géographie (France Outre-Mer, Régions) et des modes de diffusion (direct, rattrapage et vidéo à la demande). Par exemple, les programmes de la chaîne La 1ère sont réservés aux départements d'outre-mer, tandis que les vingt-quatre antennes régionales de France 3 sont accessibles en rattrapage quelque soit la région où l'on se trouve.

Il existe aussi un autre portail, Pluzz VàD, qui gère quant à lui le service de vidéo à la demande du groupe. On peut alors louer un programme pour 48 heures ou acheter le programme, pour y accéder de façon illimitée. Les programmes disponibles sont des

programmes déjà diffusés sur une des chaînes du groupe ou en avant-première : des séries, des dessins animés, des documentaires, des longs métrages, et des spectacles.

France Télévisions a aussi lancé son portail de services HbbTV fin Août 2011, qui est le seul à être multi-chaînes.

Figure 41 : Aperçu du guide des programmes de France Télévisions depuis le portail HbbTV

Le guide des programmes comprend une illustration, le titre du programme et un bref résumé, pour les programmes diffusés en direct sur toutes les chaînes du groupe, mais aussi pour les programmes suivants et les programmes de la soirée (cf. figure 41). En un « clic » depuis la télécommande, les émissions peuvent alors être « twittées » ou « facebookées », c'est-à-dire être directement commentées sur les réseaux sociaux, sans recourir à son smartphone ou son ordinateur.

Pour la météo, le téléspectateur peut entrer son code postal grâce à sa télécommande, qui sera alors enregistré, et il accède à la météo locale. Il peut aussi accéder au dernier flash météo en vidéo à la demande.

Figure 42 : Aperçu de la météo depuis le portail HbbTV

3. Comparatif des Formats Vidéo et Audio de Différentes Chaînes sur Différents Vecteurs de Diffusion

Le site http://www.digitalbitrate.com/ est un site qui présente les résultats de mesures des signaux numériques audio et vidéo des chaînes diffusées via la télévision numérique terrestre, l'ADSL, le satellite et le câble. On observe de grandes disparités selon qu'une même chaîne est transmise par la TNT ou l'ADSL ou le câble, alors que le signal original est le même.

Le 1er Avril 2013, j'ai donc décidé de faire un relevé des mesures moyennes pour les chaînes TF1, TF1 HD, France 2, France 2 HD, France 3, Arte, Arte HD, M6, M6 HD et D8, transmises par la TNT, par l'ADSL (par le fournisseur d'accès à Internet Free, qui est le seul disponible ici), par le câble (Numericable) et par le satellite (bouquet Fransat via opérateur Eutelsat). Les mesures des chaînes France 2 (SD et HD) et France 3 se trouvent en annexe A, pages 195 et 196.

Les différences de format et de débit observées entre les chaînes, et pour une même chaîne entre les différents vecteurs de diffusion expliquent les différences de qualité observées. Au niveau de l'audio, bien que les débits soient généralement les mêmes, les flux audio sont parfois compressés par les opérateurs Internet et du câble. Les opérateurs Internet proposent en plus des définitions standard et haute définition, un flux bas débit, pour les téléspectateurs qui sont très éloignés du DSLAM, et ne disposent pas d'un débit suffisant pour regarder les chaînes en définition standard. En bas débit, les signaux vidéo et audio sont très compressés.

Chapitre 4 : Le MPEG Surround : Théorie

Jusqu'à aujourd'hui, la majorité des technologies de codage d'audio multicanal réalisent un encodage discret des canaux audio, c'est-à-dire que chaque canal est codé séparément dans un flux, qui est diffusé (par exemple un flux en Dolby Digital Plus à 256 kbits/seconde en diffusion TNT en haute définition). Le décodeur du téléspectateur se charge alors de reconstruire les canaux discrets. Ce type de codage restitue de façon plutôt fidèle le signal original, aux pertes dues à la compression près, mais nécessite beaucoup de débit.

D'autres technologies utilisent des techniques de matriçage, qui mélangent les canaux audio sur les deux canaux d'un flux stéréophonique, moyennant des algorithmes précis décrivant les sommations des canaux et les rotations de phase. Le format le plus ancien est apparu dans les années 1970 : il s'agit du Dolby Surround, et plus récemment d'autres formats matricés sont nés, comme le Dolby Pro Logic II par exemple. Les principaux avantages des techniques de matriçage sont le gain de débit, ainsi que la compatibilité avec une majorité de décodeurs. Néanmoins, la qualité de restitution est loin d'être idéale.

Depuis quelques années, une nouvelle technologie est née, alternative aux codages en canaux discrets ou au matriçage, qui permet de réaliser un downmix stéréophonique du signal multicanal et d'en extraire les informations de spatialisation pour les coder dans un flux annexe au flux stéréophonique, les données de spatialisation étant peu volumineuses. Cette technologie a été développée par Coding Technologies et Fraunhofer : il s'agit du MPEG Surround. Tout comme les techniques de matriçage, ce format permet une compatibilité 5.1/stéréophonique, et même une compatibilité avec le binaural.

Le standard MPEG SAOC (Spatial Audio Object Coding) utilise cette technologie, en permettant à l'auditeur de placer un objet à sa guise dans l'espace sonore.

1. TECHNOLOGIE SAC : SPATIAL AUDIO CODING

La technologie de codage spatial de l'audio, aussi appelé SAC, est un codage qui permet d'extraire les données de spatialisation et de les coder dans un flux particulièrement compact, et de réaliser en parallèle un downmix stéréophonique compatible avec la plupart des décodeurs. Le flux de données, peu volumineux, est alors annexé au flux audio stéréophonique, dans un seul flux MPEG-4. Le décodeur réalise l'opération inverse et reconstruit un signal de haute qualité, à l'aide du downmix et des paramètres de spatialisation.

Afin d'atteindre de meilleurs taux de compression, ce codage exploite les redondances inter-canaux dans les signaux audio multicanaux.

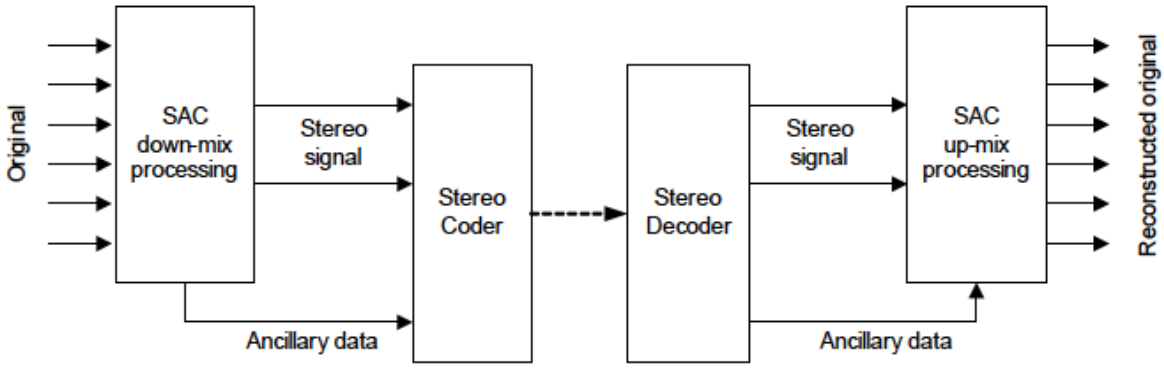


Figure 43 : Principe du codage Spatial Audio Coding

Le codage SAC est une extension de la technologie « Stéréo paramétrique » incluse dans le codec HE-AAC v2.

Les paramètres spatiaux sont issus des valeurs des différences interaurales de niveau, des différences interaurales de temps ou de phase et des degrés de cohérence

interaurale. Mais la technologie SAC ne se limite pas à ces paramètres spatiaux, elle inclut aussi des méthodes de prédiction pour modéliser les relations entre canaux.

Contrairement aux codages matricés tels que le MPEG-2, le Dolby Surround ou le Dolby Pro Logic, qui sont performants en matière de gain de débit et de compatibilité avec la majorité des récepteurs 2.0, les codages paramétriques ont une meilleure fidélité de restitution, notamment pour les signaux peu corrélés, et permettent eux aussi une importante réduction de débit.

Le codage MPEG Surround exploite cette technologie.

2. LE MPEG SURROUND

La technologie MPEG Surround a été développée par la société Fraunhofer IIS, puis normalisée par le Moving Picture Expert Group en 2007, sous le nom de ISO/IEC 23003-1. Le codage MPEG Surround utilise le principe du codage spatial de l'audio (SAC).

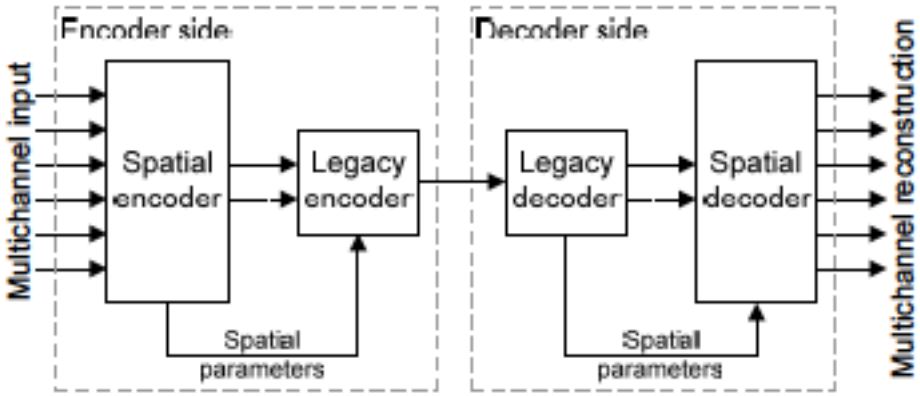
On peut considérer la technologie MPEG Surround comme une extension de la technologie Stéréo paramétrique, présente dans le codec HE-AAC v2, mais le MPEG Surround supporte davantage de débits et est de meilleure qualité.

2.1 PRINCIPE

Le codage MPEG Surround réalise un downmix stéréophonique du signal multicanal, tout en extrayant les données de spatialisation. Le downmix stéréophonique est alors encodé avec un autre codeur, souvent en HE-AAC ou en AAC, et le flux de données est codé séparément. Ces deux flux sont ensuite encapsulés dans un flux MPEG-4, et le signal est ainsi transporté.

A l'arrivée, si le décodeur est compatible MPEG Surround, il recrée un signal 5.1 à partir du downmix stéréophonique et des données de spatialisation. S'il n'est pas compatible, il lit alors le downmix stéréophonique.

Figure 44 : Schéma de principe du codage et décodage en MPEG Surround



Dans sa plus simple implémentation, le MPEG Surround permet de coder un signal 5.1 en un flux stéréophonique accompagné de données de spatialisation. Cependant ce codage est évolutif, et supporte en entrée un signal multicanal possédant jusqu'à vingt-sept canaux, et un choix de canaux pour le downmix libre, à condition que ce nombre soit inférieur au nombre de canaux du signal original.

2.2 Encodeur MPEG Surround

2.2.1. Principe d'encodage

Les informations de spatialisation sont extraites du signal multicanal original durant l'encodage, en même tant que le downmix stéréophonique est généré, puis le downmix est encodé avec un autre codeur (principalement en HE-AAC, en AAC ou en MPEG 1 Layer 2). Le flux audio et le flux de données sont alors encapsulés dans un flux MPEG-4.

Pour cela, l'encodeur MPEG Surround utilise un ensemble de procédés complexes : des fonctions de filtrage, des blocs de downmix et d'analyse, et des blocs de synthèse.

Chaque canal issu du signal multicanal subit d'abord une pré-atténuation, pour ajuster les niveaux des canaux entre eux. Le codage MPEG Surround supporte des pré-gains paramétrables par l'utilisateur, compris entre 0 et – 6 dB par pas de -1,5 dB. Le canal LFE peut lui être atténué de 0 à 20 dB, par pas de -5 dB. En sortie d'encodage, le niveau du downmix peut encore être contrôlé par l'utilisateur : il peut être atténué de 0 à 12 dB, par pas de -1,5 dB. Ces pré et post-gains sont renseignés dans le flux de données, et sont donc réversibles lors du décodage.

Après l'étape du pré-gain, les canaux subissent une analyse temporelle et sont convertis dans le domaine fréquentiel à l'aide d'une banque de filtres. Cette banque de filtres tente de simuler les résolutions spectrale et temporelle de l'oreille humaine, et extrait donc plusieurs bandes paramétriques. Le nombre de bandes paramétriques définit la finesse de restitution. Les paramètres spatiaux sont extraits lors du downmix, puis quantifiés et encodés dans un flux de données. Le downmix est ensuite reconverti dans le domaine temporel par une banque de filtres de synthèses et un post-gain est appliqué. Les données de spatialisation sont quantifiées en un seul flux, les redondances sont exploitées pour minimiser encore davantage le débit. Le downmix est ensuite encodé dans un autre codec perceptuel.

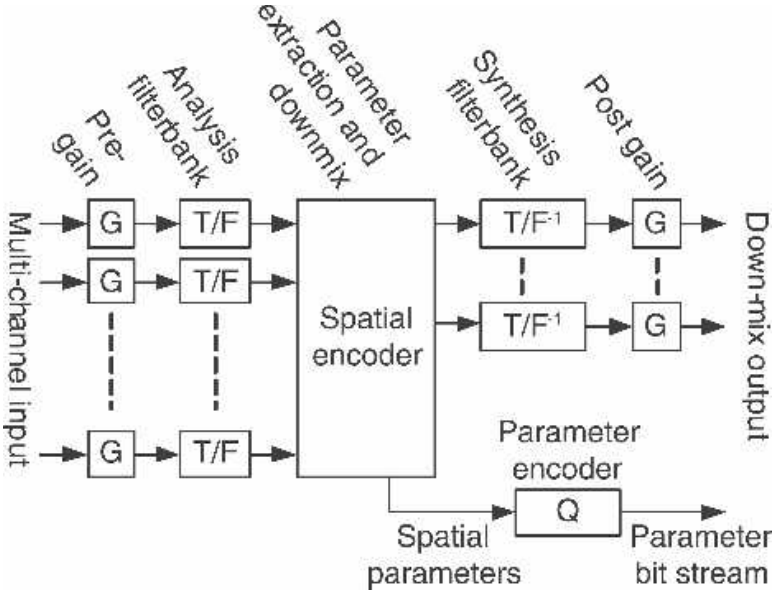


Figure 45 : Schéma de principe d'un encodeur MPEG Surround

2.2.2. STRUCTURES DE BASE

Le codage MPEG Surround permet une grande flexibilité au niveau du nombre de canaux d'entrée et de sortie, grâce à l'utilisation d'éléments simples lors du downmixing, qui peuvent être combinés à souhait pour construire des structures plus complexes.

Les deux éléments de base utilisés sont une structure appelée « One-To-Two » (abrégée en OTT, ou structure « 1 vers 2 ») et « Two-To-Three » (abrégée en TTT, ou structure « 2 vers 3 »). L'élément OTT reproduit deux canaux avec un canal en entrée et des données de spatialisation ; et l'élément TTT reproduit en sortie trois canaux avec deux canaux en entrée et des données de spatialisation. Les structures inverses « Reverse OTT » et « Reverse TTT » permettent quand à elles de réaliser des downmixings.

L'encodeur MPEG Surround extrait les trois paramètres suivants avec les éléments « R-OTT » et « R-TTT » :

♪ le CLD (Channel Level Differences) ou indice de différence de niveau inter-canal, qui décrit les différences de niveaux entre deux canaux,

♪ l'ICC (Inter Channel Correlation / Coherences) ou indice de corrélation inter-canal, qui détaille la quantité de corrélation ou de cohérence entre deux canaux, paramètre primordial pour la sensation de spatialisation,

♪ le CPC (Channel Prediction Coefficients) ou coefficients de prédiction, qui autorisent la prédiction et la reconstruction d'un troisième canal à partir de deux canaux.

Les éléments « R-OTT » et « R-TTT » génèrent un signal résiduel appelé « res », qui contient une erreur modélisée, qui permettent, en plus des paramètres spatiaux de reconstruire la forme d'onde la plus proche possible de l'original.

Ces paramètres sont codés dans le flux de données, à faible débit, et ce flux est transmis avec le downmix stéréophonique dans un flux MPEG-4 unique. Le débit du flux de données de spatialisation est à la discrétion de l'utilisateur, mais évidemment plus ce débit est élevé, meilleure sera la restitution du signal 5.1.

2.2.3. ENCODAGE D'UN SIGNAL 5.1 EN MPEG SURROUND

Afin d'encoder un signal multicanal 5.1, l'algorithme utilise trois éléments « R-OTT » : le premier réalise un downmixing des canaux avant gauche et arrière gauche, et extrait les paramètres de différence inter-canal de niveau (CLD) , de corrélation inter-canal (ICC), et un signal résiduel ; le second réalise un downmixing des canaux avant droit et arrière droit, en extrayant les mêmes paramètres, et le troisième élément réalise un downmixing des canaux central et LFE et n'extrait que des paramètres de différence inter-canal de niveau (CLD). Ces trois downmix monophoniques sont à leur tour injectés dans une cellule « R-TTT », qui réalise un downmix stéréophonique Left Only Right Only (L_0R_0), et fournit les coefficients de prédiction, les différences de niveau inter-canal, la corrélation inter-canal, et un signal résiduel. Tous ces paramètres peuvent être compactés et occupent peu de bande passante.

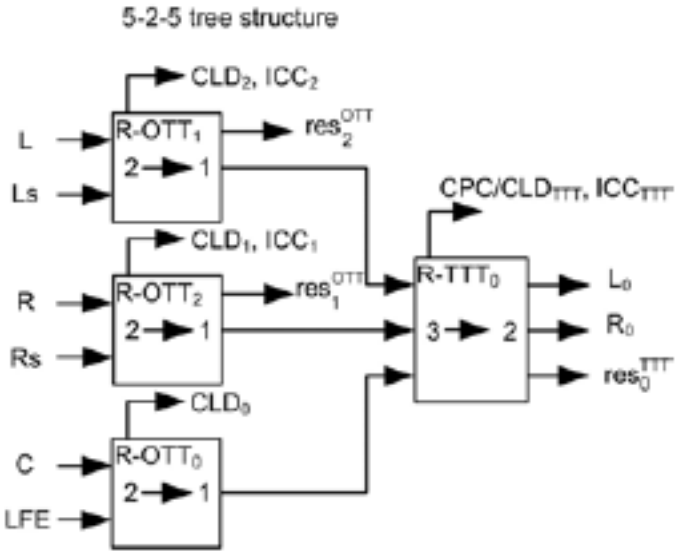


Figure 46 : Encodage MPEG Surround : extraction des paramètres de spatialisation et downmixing

2.2.4. MODES D'ENCODAGE : MODE INTERNE ET MODE EXTERNE

Il existe deux modes d'encodage : un mode dit « Internal Downmix Mode » et un mode appelé « External Downmix Mode », ce dernier ayant été créé afin de se rapprocher des conditions de production où un ingénieur du son produit souvent un programme stéréophonique et un programme multicanal.

Le mode « Internal Downmix » fournit, de façon automatique, un downmix standard et un flux de données de spatialisation, comme décrit ci-dessus.

Au contraire le mode « External Downmix » est un mode artistique, qui permet d'injecter une stéréo externe, produite par l'ingénieur du son. Son schéma de principe est illustré sur la figure 47. L'encodeur MPEG Surround crée son downmix automatique, et calcule les différences entre le downmix « artistique » externe et le downmix interne qu'il a fabriqué. Ces différences sont alors codées dans les données de spatialisation, de manière paramétrique pour les applications nécessitant un faible débit, ou par un signal codant la différence de forme d'onde, comme codage résiduel. Ce mode requiert une parfaite synchronisation des signaux d'entrée.

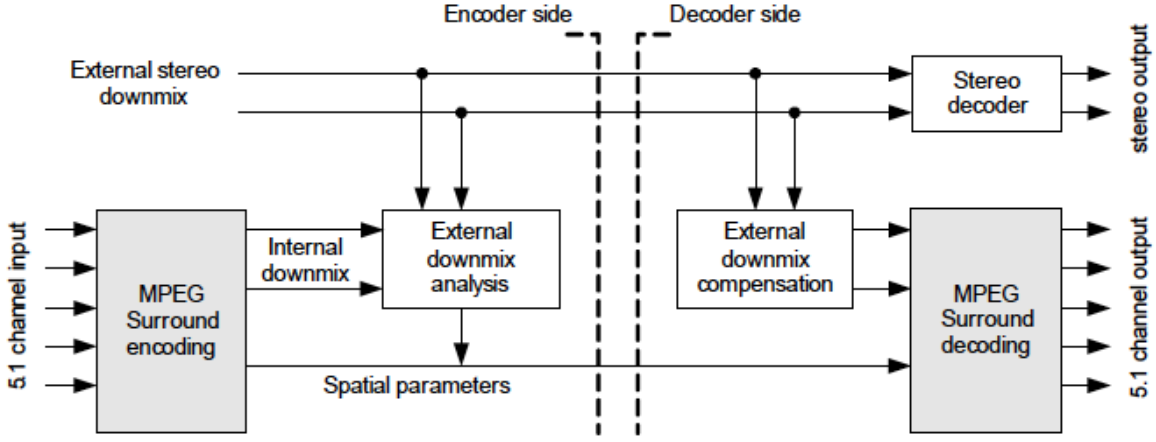


Figure 47 : Schéma de principe du mode "External Downmix"

Afin de garantir une parfaite rétro-compatibilité avec des systèmes existants, le codage MPEG Surround propose aussi un encodage matricé du downmix stéréophonique, qui pourra être décodé par des systèmes matricés multicanaux. Le décodeur MPEG Surround dématrice le downmix préalablement matricé et reconstruit le signal multicanal à partir du downmix et des paramètres spatiaux, sans dégradation du signal, grâce au dématriçage intégré.

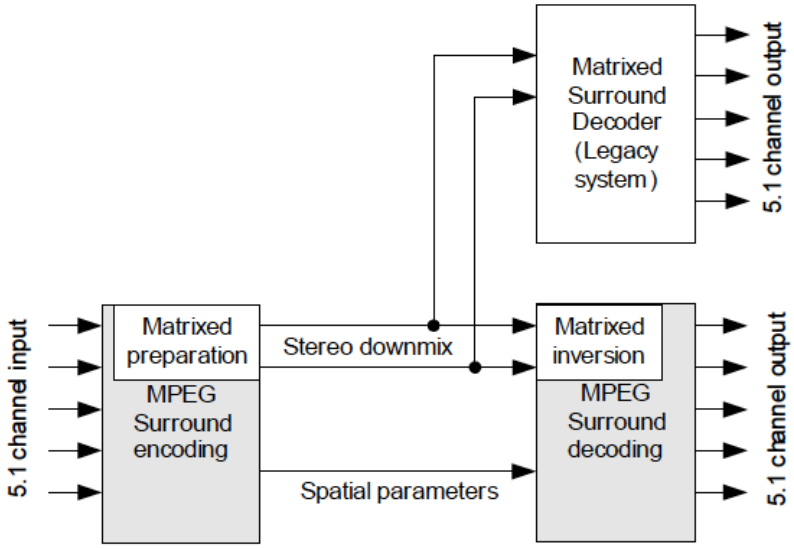


Figure 48 : Compatibilité du downmix MPEG Surround avec un décodeur matricé

2.2.5. DÉBIT

Les données de spatialisation ne requièrent pas beaucoup de débit (environ 10%), l'essentiel du débit étant dédié codeur principal. Le MPEG Surround supporte des données de spatialisation encodées à un débit allant de 3 kbits/seconde à 32 kbits/seconde. Néanmoins, si on a besoin d'une grande transparence et d'une qualité optimale, il faut disposer du plus haut débit possible.

La figure 49 montre que le format MPEG Surround est meilleur que les technologies de matriçage, encore couramment utilisées.

De plus, si on choisit une grande résolution fréquentielle, c'est-à-dire des bandes de fréquences étroites et nombreuses, on assure une meilleure qualité puisque les éléments audio seront bien détaillés, mais cela nécessite beaucoup de débit. Il en va de même pour la résolution temporelle.

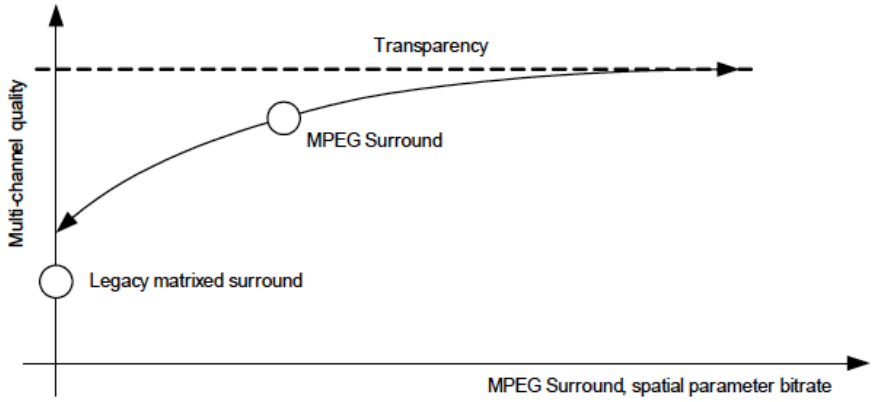


Figure 49 : Qualité du MPEG Surround en fonction du débit

2.3 DÉCODEUR MPEG SURROUND

Le décodeur MPEG Surround décode d'abord le downmix stéréophonique, puis reconstruit le signal multicanal en exploitant les données auxiliaires.

Le décodeur peut fonctionner dans le domaine du temps ou dans le domaine fréquentiel (QMF).

2.3.1. FILTRES QMF

Les filtres miroirs en quadrature, abrégés en QMF, permettent de reconstruire des signaux à partir de composantes basses fréquences et hautes fréquences, après les avoir sous-échantillonnées. Cet échantillonnage ne respecte pas le théorème de Shannon, à savoir $F_{échantillonnage} > 2*F_{maximale}$, le signal est donc sous-échantillonné et l'on observe un repliement spectral. Néanmoins, les filtres passe-bas et passe-haut sont complémentaires et garantissent une reconstruction exacte. Ces filtres respectent les conditions d'application du deuxième théorème de Nyquist qui concerne l'échantillonnage de

signaux numériques : on peut transmettre un signal sur un canal très bruité, avec peu d'erreurs, à condition de ne pas dépasser un certain débit, qui est calculable.

La banque de filtres QMF utilisée est identique à celle utilisée dans le codage High-Efficiency Advanced Audio Coding, qui sert à améliorer le taux de compression avec la technologie Spectral Band Replication. C'est pour cette raison entre autres que le codec MPEG Surround semble très performant quand il est associé au codeur principal HE-AAC car il évite une étape de transcodage.

2.3.2. Décodage d'un Signal Stéréophonique vers un Signal 5.1

Si aucun signal résiduel n'a été transmis, il faut générer un signal pour garantir une reconstruction correcte, qui soit indépendant du signal d'entrée mais qui ait un spectre et un timbre proches de celui-ci. Ces signaux sont alors générés par des filtres de décorrélation.

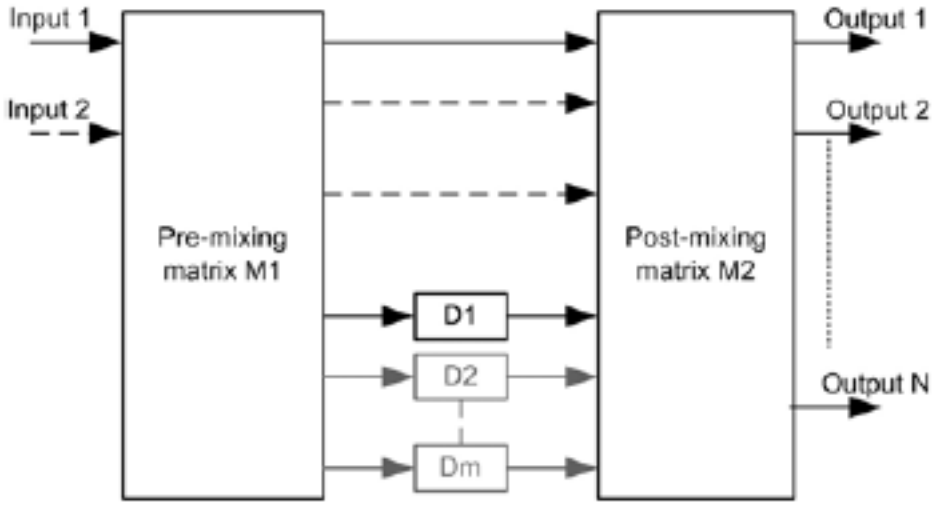


Figure 50 : Décodage MPEG Surround

Dans ce cas, un pré-dé-matriçage a lieu, avant d'ajouter l'étape de décorrélation, puis le véritable dé-matriçage se fait et un signal multicanal est reconstruit.

2.3.3. Principe du Décodage

Avant décodage, un pré-gain est appliqué au downmix stéréophonique. Le signal est ensuite converti dans le domaine des fréquences, à l'aide de filtres QMF, puis le décodeur reconstruit chaque canal avec les données spatiales et les informations du downmix stéréophonique. Les canaux reconstruits subissent à nouveau une analyse et sont reconvertis dans le domaine du temps par des blocs de synthèse. Un post-gain est appliqué juste avant la sortie, égal à l'inverse du pré-gain appliqué avant l'encodage.

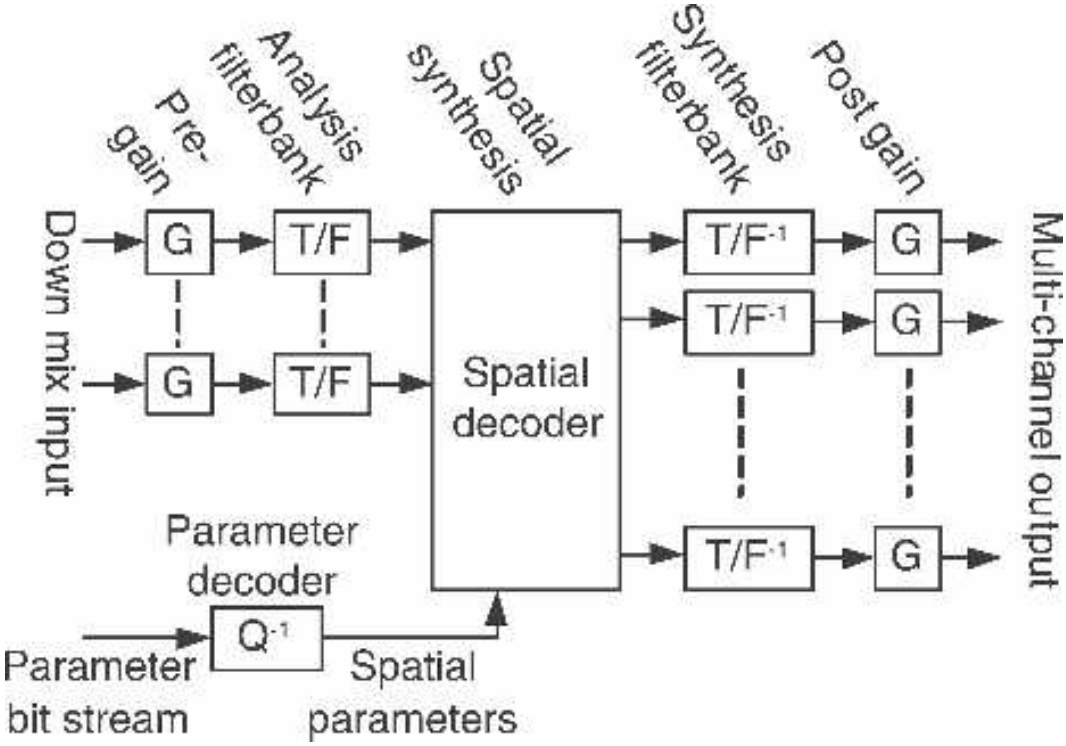


Figure 51 : Schéma de principe d'un décodeur MPEG Surround

2.3.4. Structure de Décodage

Le décodeur MPEG Surround reçoit le downmix L_0R_0, ainsi qu'un signal résiduel et les paramètres de spatialisation (CPC/CLD : différence interaurale de phase et de niveau, ICC : degré de cohérence). A l'aide d'un module TTT, trois signaux monophoniques sont créés (Gauche total, Droite totale et centre) puis ils transitent tous dans un élément OTT, qui permet de reconstruire les six canaux originaux.

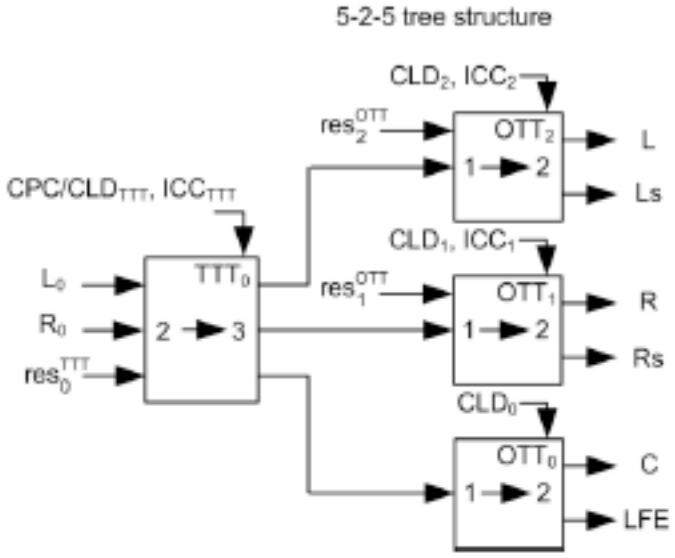


Figure 52 : Structure de décodage MPEG Surround : reconstruction du signal 5.1 à partir du downmix et des données de spatialisation

2.3.5. Modes de Décodage : le Mode Normal et le Mode Amélioré

Pour le décodage, il existe deux modes : un mode appelé « Normal Mode » et un mode nommé « Enhanced Matrix Mode », que l'on peut traduire par mode matriciel amélioré.

Le mode « Normal » est choisi quand le flux à décoder possède des données de spatialisation extraites du signal multicanal original.

Si aucun flux de données de spatialisation n'avait été extrait du signal original, le mode « Enhanced Matrix Mode » permet au décodeur d'estimer les paramètres de spatialisation à partir du downmix, avec des blocs d'analyse constitués de filtres QMF (cf. figure 53).

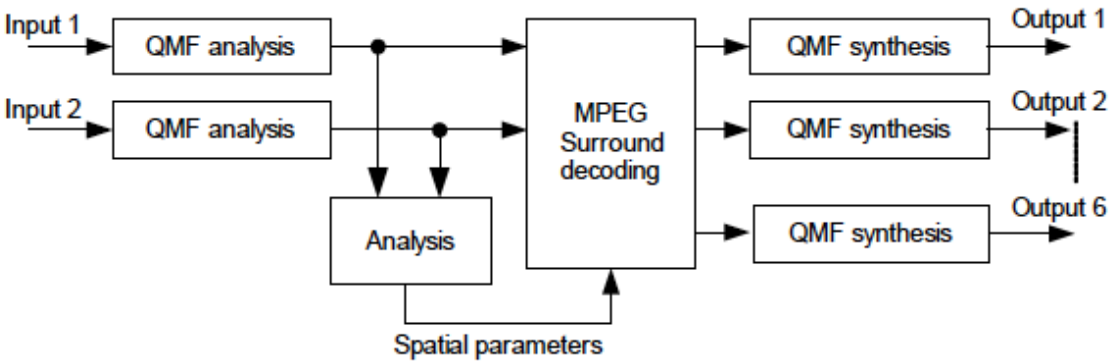


Figure 53 : Schéma de principe du décodeur en mode matriciel amélioré

2.3.6. COMPATIBILITÉ BINAURALE

De plus, le MPEG Surround est parfaitement compatible en binaural. Une fois le signal multicanal encodé, le flux contenant le downmix stéréophonique et les données auxiliaires est alors transmis. En mode binaural, le décodeur possède un encodage binaural, qui est capable de décoder le flux MPEG-4 et de construire le signal binaural correspondant aux HRTF insérées et simulant le signal multicanal au casque, sans reconstruire les six canaux de façon discrète. Cette option économise alors des ressources processeur.

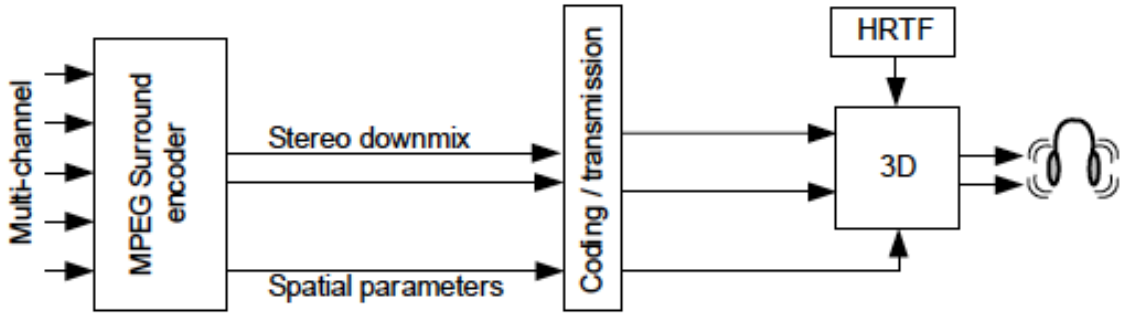


Figure 54 : Schéma du décodeur binaural MPEG Surround

3. Essais Préalablement Réalisés par d'Autres Laboratoires

Voici un exemple de tests perceptifs menés selon la méthodologie MUSHRA[23], dans trois laboratoires différents, avec treize participants considérés comme des auditeurs experts. Les résultats de ces tests ont été publiés dans la 123ème convention de l'AES[24], à New-York, en octobre 200. Dix extraits sonores ont été utilisés : des applaudissements, un extrait de la Toccata en Ré mineur de Bach, jouée sur un orgue dans une église, un extrait d'un orchestre brass-band (orchestre constitué d'instruments à percussions et de cuivres), un extrait avec des transitoires de guitare, un extrait de clavecin, un extrait instrumental de l'opéra Lohengrin, un extrait d'harmonica, un extrait d'un chœur, un extrait d'une dramatique radio et un extrait de trompette. Les auditeurs devaient évaluer, en référence au signal original, une référence cachée, une ancre, c'est-à-dire le signal original, auquel on avait appliqué un filtre passe-bas à 3,5 kHz, un signal 5.1 encodé en six canaux discrets en AAC à 320 kbits/seconde, le codec MPEG-1 Layer-2 à 256 kbits/seconde combiné au Dolby Prologic II, le codec HE-AAC en canaux discrets à 64 et 160 kbits/seconde, le codec MPEG Surround combiné au codeur HE-AAC à 64, 96 et 160 kbits/seconde, le codec MPEG Surround combiné au codeur MPEG-1 Layer-2 à 192 et 256 kbits/seconde, et enfin le codec MPEG Surround combiné au codeur AAC-LC à 192 kbits/seconde. Les participants doivent comparer les codecs sur ces différents extraits, et leur attribuer une note comprise entre 0 et 100, 0 étant un signal très dégradé et 100 un signal d'excellente qualité, équivalente à l'original.

La moyenne obtenue par les références cachées est très proche de 100, quelque soient les extraits. Les ancres, signaux filtrés à 3,5 kHz, obtiennent les notes les plus faibles, entre 15 et 22. La configuration avec le codec MPEG-1 Layer-2 à 256

[23] La méthodologie MUSHRA, décrite dans la recommandation ITU-R BS.1534, sera explicitée dans le chapitre 5, 3.1. Page 114.

[24] Etude publiée dans l'article « A study of MPEG Surround quality versus bit-rate », Proceedings of the 123rd AES Convention, New York, NY, USA, 2007 October 5-8.

kbits/seconde combiné au Dolby Prologic II obtient une note moyenne de 45/100. L'encodage en canaux discrets d'un signal multicanal en AAC à 320 kbits/seconde est de très grande qualité, et rafle une moyenne de 96/100.

Concernant le MPEG Surround, combiné au codec HE-AAC, on observe une amélioration de la qualité croissante avec le débit. A 64 kbits/seconde, ce codec obtient une moyenne de 72,6/100, ce qui est déjà une bonne note. A 96 kbits/seconde, il remporte une moyenne de 80,6/100, tandis qu'à 160 kbits/seconde, il rafle la note de 89,5/100. On peut d'ailleurs noter que l'encodage en canaux discrets en HE-AAC à 160 kbits/seconde ou l'encodage en MPEG Surround combiné au codeur HE-AAC à 160 kbits/seconde, ou l'encodage en MPEG Surround combiné au codeur AAC-LC à 192 kbits/seconde obtiennent une moyenne semblable : ils apparaissent donc de qualité équivalente. Au contraire, le MPEG Surround combiné au MPEG-1 Layer 2 semble nécessiter davantage de débit tout en étant un peu moins performant : une moyenne de 73,7/100 à 192 kbits/seconde, et une moyenne de 85/100 environ à 256 kbits/seconde. La combinaison du MPEG-1 Layer-2 au Dolby ProLogic obtient une moyenne faible, inférieure à 50/100. La figure 55 présente les notes moyennes et l'intervalle de confiance à 95% obtenus par chaque codec, pour chaque extrait. La figure 56 est un graphique qui montre les moyennes globales obtenues par chaque codec, selon le débit, tous extraits confondus.

On observe alors que l'ajout du MPEG Surround améliore considérablement la qualité par rapport à un encodage en canaux discrets en HE-AAC à 64 kbits, même pour des signaux complexes tels que des applaudissements, toutefois cette différence de qualité se réduit avec l'augmentation du débit.

Cette étude conclut donc que le MPEG Surround est très performant, il permet des réductions de débit intéressantes, tout en conservant une bonne qualité, et pourrait donc être utilisé même dans des domaines audio qui requièrent une excellente qualité.

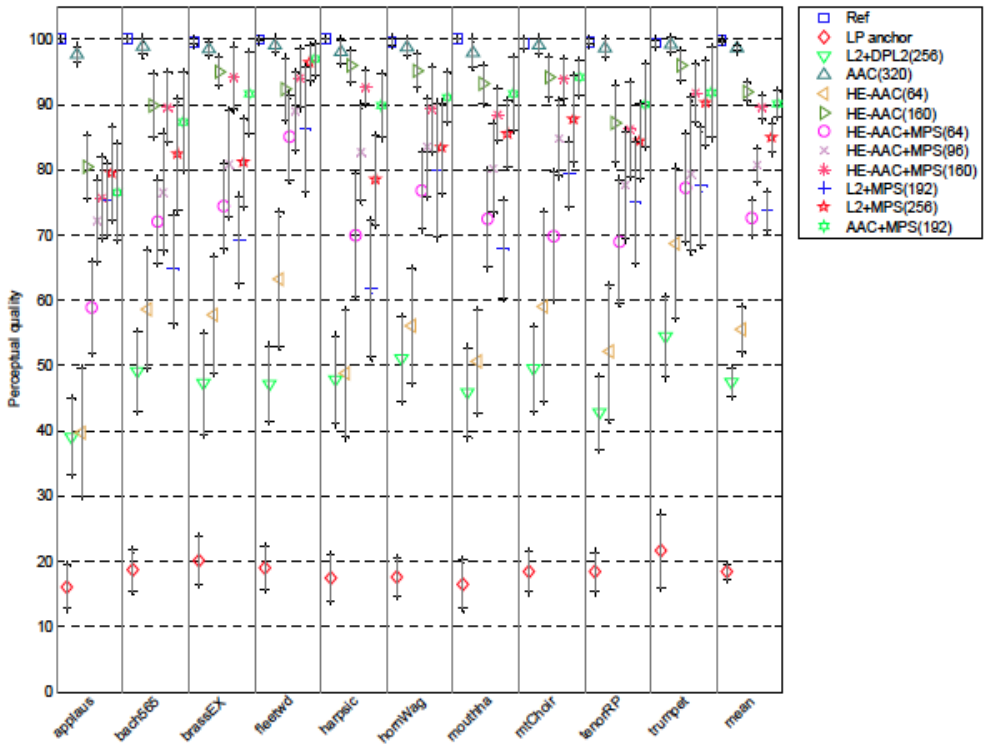


Figure 10 - Listening test results for various excerpts (abscissa) and codecs (symbols). Errorbars denote the 95% confidence interval of the mean.

Figure 55 : Diagramme des résultats des tests perceptifs, selon les extraits en abscisses et les notes des différents codecs en ordonnée (les barres d'erreur montrent un intervalle de confiance à 95%)

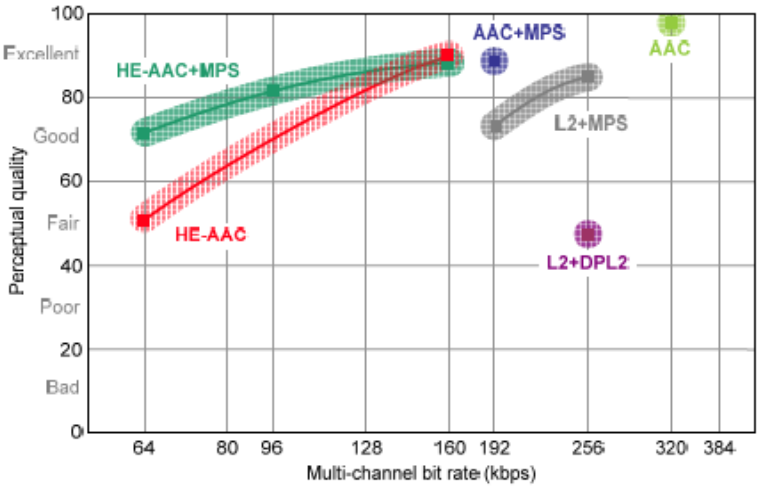


Figure 11 - Perceptual quality (averaged across excerpts) as a function of bit rate for various coder configurations.

Figure 56 : Diagramme de la qualité moyenne perçue en fonction du débit des différents codecs

Ce test perceptif présente l'avantage d'une comparaison d'un grand nombre de codecs, mais nécessite beaucoup de temps pour mener à bien ce test. En revanche, les extraits sont essentiellement des extraits musicaux, or ce type de programme n'est pas représentatif de la majorité des programmes télévisés diffusés en 5.1.
Je vais donc réaliser un test perceptif avec moins de codecs, mais de véritables extraits de programmes télévisés.

4.0 L'Implémentation du MPEG Surround en 2013

Fraunhofer propose aujourd'hui des implémentations du MPEG Surround pour de l'électronique embarquée (Set Top Boxes, récepteurs portables, etc.), ainsi que pour des stations de travail audionumériques. Ce codec nécessite l'achat d'une licence, ce qui est aujourd'hui un frein à son intégration dans les équipements grands publics. Bien qu'il ne dispose pas de l'exclusivité de ce standard, Fraunhofer est aujourd'hui le seul fabricant de logiciels d'encodage en MPEG Surround.

Sonnox a implémenté le codec MPEG Surround au sein du plug-in Fraunhofer ProCodec, plugin incluant différents codeurs et décodeurs audio, parmi lesquels on retrouve les codecs AAC-LC (Advanced Audio Coding, Low Complexity), HE-AAC et HE-AAC v2 (High Efficiency – Advanced Audio Coding), MPEG 1 Layer 3 (mp3), mp3 Surround, mp3 HD, et évidemment le codec MPEG Surround.

4.1. Compatibilité avec les Lecteurs Grands Publics

L'un des avantages du MPEG Surround réside dans sa parfaite compatibilité stéréophonique/multicanal.

Si le téléspectateur ne dispose pas de décodeur MPEG Surround, ni d'équipement home-cinéma, son lecteur compatible MPEG-4 doit alors lire l'audio stéréophonique. Au contraire, si le téléspectateur dispose d'un décodeur MPEG Surround et d'un équipement de reproduction audio multicanal, son lecteur décode alors le flux MPEG-4 et délivre un signal audio 5.1.

Il était donc primordial de s'assurer de la compatibilité stéréophonique du MPEG Surround avec les lecteurs multimédias grand public actuels. J'en ai donc testé plusieurs, les lecteurs suivants lisent parfaitement le flux MPEG-4, avec l'extension .m4a :

- ♪ QuickTime Player, version 10.2 (603.12) sur Mac OS X 10.8.3 et version 7.7.3 (1680.64) sur Windows 7 SP1,
- ♪ VLC Media Player, version 2.0.5 sur Mac OS X 10.8.3 et sur Windows 7 SP1,
- ♪ Real Player, version 12.0.1 (1750) sur Mac OS X 10.8.3 et version 16.0.1.18 sur Windows 7 SP1,
- ♪ Winamp Sync Bêta, version 0.8.1 sur Mac OS X 10.8.3,
- ♪ Winamp, version 5.63 sur Windows 7 SP1,
- ♪ Lecteur Windows Media, version 12.0.7601.17514 sur Windows 7 SP1,
- ♪ Flip Player version 3.1.0.24 (98969) sur Mac OS X 10.8.3

En revanche, iTunes, version 11.0.2.26 sur Mac OS X 10.8.3 et sur Windows 7 SP1, refuse d'importer un fichier audio avec l'extension .m4a. Il faut donc modifier l'extension des fichiers et remplacer .m4a par .mp4. Avec cette dernière, iTunes accepte d'importer les fichiers MPEG-4 et les lit parfaitement.

J'ai aussi essayé de les importer sur des stations audionumériques, comme ProTools 10 et Samplitude 11, même si le MPEG Surround n'a pas vocation à être travaillé en post-production. Ces deux stations nous proposent de convertir le fichier MPEG-4, qu'elles détectent comme ayant une fréquence d'échantillonnage de 24 kHz, à la bonne fréquence de la session, à savoir 48 kHz. Quand elles convertissent ces fichiers, on obtient un fichier audio avec un son une octave en dessous. ProTools convertit un fichier de la bonne durée, mais puisque le son est transposé une octave en-dessous, la deuxième moitié du fichier audio est coupée. Quant à Samplitude, il fabrique un fichier d'une durée double par rapport à l'original. Sur Samplitude, pour pallier à ce problème, il faut préciser lors de l'importation qu'on ne veut pas de conversion, le logiciel prévient alors que le fichier sera lu à la mauvaise vitesse. Mais en fait, c'est le seul moyen de lire

correctement le fichier MPEG-4, d'une durée proche de l'original, et dans la bonne tonalité. Par contre, si on a pu observer que le fichier résultant d'un encodage puis d'un décodage en MPEG Surround avait le début tronqué, le fichier en MPEG-4 a quant à lui un début allongé mais une fin tronquée.

Ces problèmes d'importation dans les séquenceurs ont plusieurs explications. Le codage HE-AAC utilise la technologie SBR (spectral band replication ou reconstruction de bande spectrale), qui travaille à une fréquence d'échantillonnage divisée par deux, puisque seul le bas du spectre est codé. Mais ces séquenceurs ne décodent pas les métadonnées et le décodage du flux HE-AAC n'est donc pas correct (l'audio est transposé une octave en-dessous, il manque la deuxième moitié du fichier, etc.).

4.2. ÉVOLUTIONS

En théorie, ce codage paraît très performant. Deux séries de tests perceptifs vont être menées, afin de déterminer quel codec est le meilleur.

Deux bémols sont quand même à signaler : le MPEG Surround est soumis à des coûts de licence, et il est inconnu du grand public pour l'instant.

CHAPITRE 5 : LE MPEG SURROUND EN PRATIQUE

Divers essais pour évaluer les qualités et les défauts du MPEG Surround ont déjà été menés par différentes personnes et les résultats ont été publiés dans le journal de l'AES[25]. Néanmoins, dans le cadre de ce mémoire, il était primordial que je réalise mes propres tests, avec mes propres programmes, afin de tirer des conclusions.

J'ai eu la chance d'accéder au laboratoire Innovations & Développement de France Télévisions, récemment mis en place et conçu en respectant les diverses recommandations, afin de mener des tests officiels. J'ai aussi bénéficié gratuitement de la licence du plugin Fraunhofer Pro-Codec de Sonnox, licence dont le laboratoire disposait, pour réaliser mes encodages et décodages en MPEG Surround.

Avec la collaboration de mes directeurs de mémoire et un enseignant, j'ai réuni quatre programmes télévisés. Des tests préliminaires m'ont ensuite permis d'affiner le protocole de tests final.

[25] AES : Audio Engineering Society, association internationale qui regroupe des professionnels de l'audio. Elle a été créée en 1948 et elle est basée à New-York. Elle publie un journal mensuel et organise plusieurs conférences par an.

1. ESSAIS PRÉLIMINAIRES

1.1. SÉLECTION DES EXTRAITS

Afin de tester et d'évaluer le codage MPEG Surround, j'ai d'abord effectué des tests préliminaires.

Je disposais de quatre programmes différents : le téléfilm *Jusqu'à l'enfer* [26] de Denis Malleval, réalisé en 2009 ; le documentaire *La vie moderne* [27] de Raymond Depardon, réalisé en 2008 ; l'opéra *L'Italiana in Algeri* [28] de Gioacchino Rossini, filmé et enregistré à l'Opéra de Nancy en février 2012, interprété par l'Orchestre symphonique et lyrique de Nancy, sous la direction de Paolo Olmi, avec le Chœur de l'Opéra National de Lorraine et le Chœur de l'Opéra-Théâtre de Metz Métropole, et enfin la version française d'un épisode de la série *U.S. Marshals : Protection des témoins*, épisode numéro six de la saison trois, intitulé *Autre temps, autres mœurs*.

Tout d'abord, j'ai visionné attentivement ces programmes avec ProTools, afin de choisir plusieurs passages intéressants et j'ai relevé les time-codes. Les vidéos étant dans des formats différents, on a décidé de les importer dans le logiciel Final Cut Pro 7, pour sélectionner les extraits, puis les exporter dans le même format HD, afin d'assurer une lecture homogène dans ProTools, malgré une perte de qualité due au transcodage.

Dans Final Cut Pro, j'ai créé un projet, avec cinq séquences aux paramètres identiques, une par programme et une pour une vidéo Syncheck, vidéo composée de flashs synchronisés avec des bips sonores à 3 kHz, qui, une fois importée dans ProTools,

[26] Le téléfilm *Jusqu'à l'enfer* de Denis Malleval m'a été fourni par M. Claude GAZEAU, avec l'aimable autorisation de la production Neyrac Films.

[27] Le documentaire *La vie moderne* de Raymond Depardon, ainsi que l'épisode de la série *U.S. Marshals : Protection de témoins*, m'ont été fournis par M. Manuel NAUDIN.

[28] L'opéra filmé *L'Italiana in Algeri*, enregistré à l'Opéra de Nancy m'a été fourni par M. Jean CHATAURET, avec l'aimable autorisation de la production François Roussillon & Associés.

permet de régler l'offset de synchronisation vidéo, à l'aide d'un boîtier de mesures. Chaque séquence comportait six pistes audio monophoniques, dans l'ordre suivant : Gauche, Droite, Centre, LFE[29], Arrière gauche, Arrière Droit. Après avoir synchronisé image et son pour chaque séquence, j'ai alors sélectionné des extraits de trente à soixante secondes, que j'ai ensuite exportés dans le même format, avec image et son inclus dans le même fichier. Ces extraits seront raccourcis dans ProTools, de façon à privilégier l'audio.

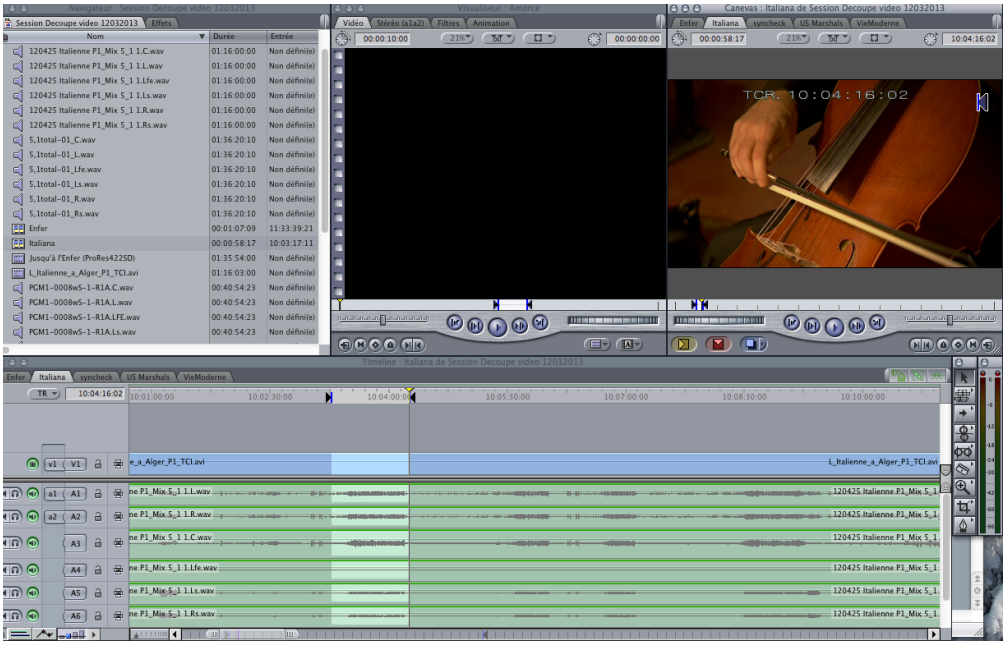


Figure 57 : Sélection d'extraits sous Final Cut Pro 7

Les images des extraits exportés seront en format HD AppleProRes 422 HQ (High Quality), avec une résolution de 1920*1080, en mode HDTV 1080i (16 :9), en pixels carrés, avec vingt-cinq images par seconde, et l'audio en format BWF[30] en canaux discrets, soit six canaux, à une fréquence d'échantillonnage de 48 kHz et une résolution de 24 bits.

[29] LFE : Low Frequency Effects, canal audio réservé aux très basses fréquences.

[30] BWF : Broadcast Wave Format est une extension du format de fichier WAVE de Microsoft, c'est un format de fichier audio multi-pistes non compressé, codé en PCM linéaire (modulation d'impulsion codée) qui contient des métadonnées telles que le time-code.

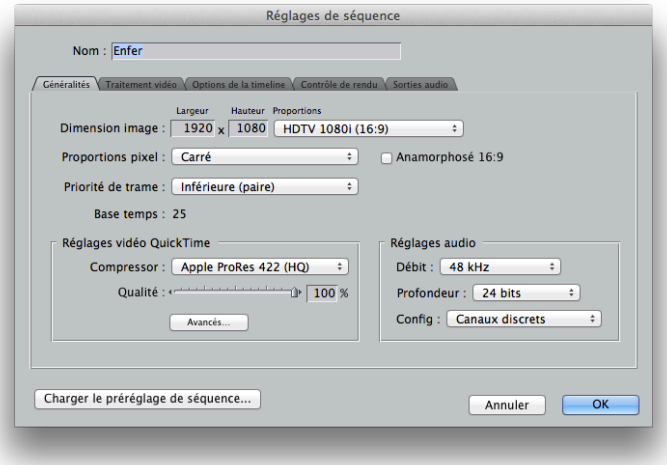


Figure 58 : Fenêtre des réglages de séquence sous Final Cut Pro

J'ai choisi six extraits d'opéra, neuf extraits du téléfilm, trois extraits du documentaire et deux extraits de la série en version française, et parmi ceux-ci je ne dois en conserver que sept[31] pour mon expérience.

1.2. Prise en Main du Plugin Fraunhofer Pro-Codec de Sonnox

Afin de préparer ma première session de tests et me familiariser avec le plugin Fraunhofer Pro-Codec, j'ai réalisé différents tests. Ce plugin est, à ce jour, le seul programme qui propose un encodeur et un décodeur MPEG Surround, bien que Fraunhofer ne dispose d'aucune exclusivité sur ce standard.

Ce plugin possède trois modes de fonctionnement : un mode appelé « Online », qui permet de comparer en temps réel sur la piste où il est inséré jusqu'à cinq codecs différents, et même d'enregistrer parallèlement la piste dans ces différents codages, à condition de bénéficier de ressources informatiques suffisantes ; un mode « Offline encode » , qui est utilisé pour encoder un fichier importé avec le codage souhaité ; et enfin

[31] Je détaillerai le choix du nombre d'extraits dans le paragraphe 3.1.3. de ce chapitre page 116.

le mode « Offline decode » qui permet, quant à lui, de décoder n'importe quel fichier importé, qui aurait été encodé dans un codage supporté par le plugin.

Ce plugin dispose des codecs avec perte suivants : mp3, mp3 Surround, AAC-LC (Advanced Audio Coding Low Complexity), HE-AAC (High Efficiency Advanced Audio Coding), HE-AAC v2, MPEG Surround, Apple AAC ; ainsi que de ces deux codecs sans perte mp3 HD and HD-AAC.

Concernant le codec MPEG Surround, ce plugin permet seulement d'encoder un signal 5.1 en un flux MPEG-4, dont le flux de données de spatialisation est encodé de 10 à 50 kbits/seconde (le débit dépend du codeur fondamental, du débit total, ainsi que du contenu audio du signal) et le downmix stéréo est ré-encodé dans un codeur fondamental : soit en HE-AAC, avec quatre débits au choix (48 kbits/seconde, 64 kbits/seconde, 96 kbits/seconde et 128 kbits/seconde), soit en AAC avec quatre débits au choix (160 kbits/seconde, 192 kbits/seconde, 256 kbits/seconde et 320 kbits/seconde). Bien que ces codecs supportent des fréquences d'échantillonnage de 44,1 kHz et 48 kHz, le fabricant préconise une fréquence d'échantillonnage de 44,1 kHz. Néanmoins, puisque je vais réaliser des tests avec image, il est préférable que mes extraits et ma session ProTools soient à la fréquence de 48 kHz, pour éviter tout problème.

Figure 59 : Fenêtre du plugin Fraunhofer Pro-Codec : Mode Online Encode

Afin de découvrir les qualités et les inconvénients du MPEG Surround, j'ai donc réalisé des essais informels avec des extraits audio du téléfilm et de l'opéra en mode « Online » : même au plus faible débit, à savoir en HE-AAC à 64 kbits/seconde, le résultat est plutôt impressionnant : la spatialisation est assez bien reproduite, même si le timbre est évidemment dégradé, notamment dans le haut du spectre. Néanmoins, quand on augmente le débit, les artefacts diminuent, mais se pose alors la question cruciale : quels sont les intérêts du MPEG Surround si le débit est équivalent au débit actuel utilisé en Dolby Digital Plus, à savoir 256 kbits/seconde ? Il faudra donc trouver le débit le plus pertinent, qui permette à la fois une économie de débit, tout en garantissant une certaine qualité.

Puisqu'une étude publiée lors de la 123ème convention de l'AES avait comparé beaucoup de codecs[32] (à savoir HE-AAC et AAC en canaux discrets ou combinés au MPEG Surround, mais aussi MPEG-1 Layer-2 combiné au Dolby Prologic II ou combiné au MPEG Surround), on a décidé d'affiner la comparaison entre trois débits différents du MPEG-Surround, combiné au codeur fondamental HE-AAC, à savoir : 64 kbits/seconde, 96 kbits/seconde et 128 kbits/seconde, avec des extraits de programmes télévisés variés (téléfilm, documentaire et opéra). J'aurais voulu aussi comparer les combinaisons MPEG Surround – HE-AAC et MPEG Surround – AAC, mais les débits proposés sont différents, et la combinaison MPEG Surround – AAC a un inconvénient majeur, malgré ses hautes performances : elle nécessite des débits assez élevés (160 à 320 kbits/seconde) : l'intérêt pour une application à la diffusion télévisuelle est donc moindre.

Dans un premier temps, on va donc se concentrer sur la combinaison MPEG Surround – HE-AAC à trois débits différents.

J'ai ensuite testé différents paramètres de l'encodeur, et notamment l'option du « trim » : soit on laisse le fichier encodé tel quel, soit on règle soit même la valeur du trim

[32] cf. Chapitre 4, 3. page 90.

lors des essais en mode « online », et elles seront alors appliquées si on choisit l'option « Use Online Trim Values » ; soit on applique un genre de limiteur, pour prévenir des saturations éventuelles au décodage, ou encore on normalise le fichier au décodage. Pour finir, les différences de niveaux entre fichier original et fichiers encodés puis décodés sont acceptables, et il était préférable de ne pas agir sur le niveau.

L'option « Output Dir is from Exports Settings », cochée sur la capture d'écran précédente, permet d'enregistrer le fichier encodé au chemin indiqué dans la fenêtre Export Settings, de plus on peut paramétrer automatiquement le nom du fichier. Par exemple, sur la figure 60 j'ai choisi le dossier dans lequel je veux enregistrer les exports, et mes fichiers seront nommés avec le nom du fichier original, suivi du débit d'encodage.

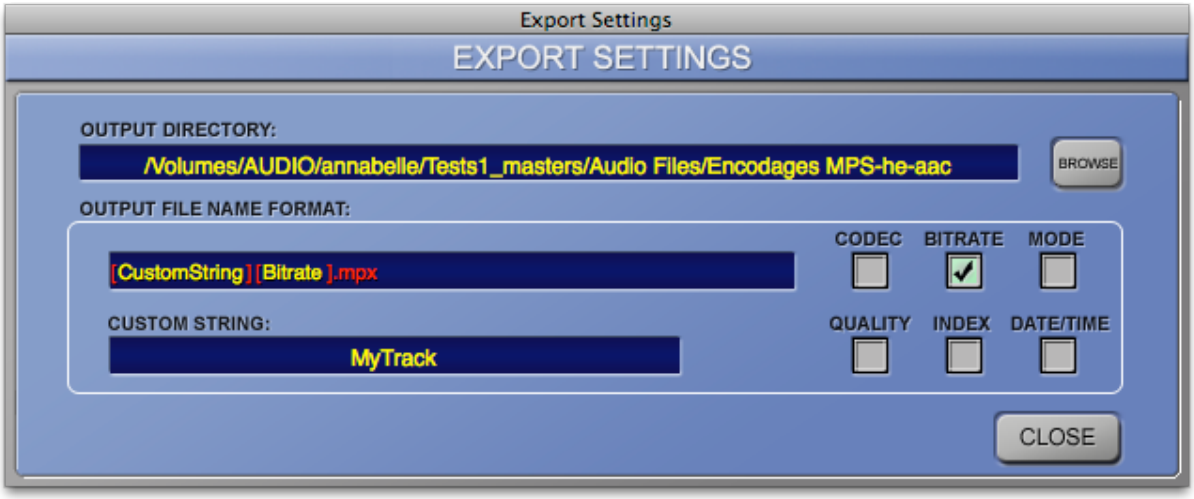


Figure 60 : Plugin Fraunhofer Pro Codec : réglages des paramètres d'exports

Dans le mode « Offline encode », on va chercher le fichier original en 5.1 entrelacé, puis on règle les paramètres d'encodage, avant de lancer l'encodage.

De la même façon, dans le mode « Offline decode », on importe le flux MPEG-4, issu de l'encodage en MPEG Surround, on choisit les paramètres de décodage (résolution en 16 ou 24 bits, et format de fichier en .wav ou .aiff), puis on lance le décodage.

Une fois familiarisée avec le plugin et les paramètres d'encodage, j'ai alors lancé les encodages, puis les décodages de tous les extraits choisis pour mon test. Hélas, au

moment de réimporter les fichiers décodés en 5.1, j'ai rencontré un premier problème, plutôt ennuyeux : les fichiers originaux et les fichiers issus du décodage du flux MPEG-4 n'ont pas la même durée, et ne sont donc pas synchrones si on synchronise le début des fichiers. Après observation, je découvre que les premières millisecondes du fichier décodé ont été coupées, quelques échantillons sont ajoutés à la fin du fichier, mais au final les fichiers décodés sont quand même plus courts que les fichiers originaux. Bizarrement, cette partie de fichier tronquée ne dépend ni de la durée du fichier (toujours inférieure à la durée d'une image, soit moins de 40 millisecondes), ni du débit d'encodage (la durée tronquée est la même à tous les débits).

Afin de contourner cet artefact d'encodage, j'ai placé un bip de synchronisation d'une fréquence 3 kHz, d'une durée d'une image, suivi de vingt-quatre images de silence, en tête de fichier, une seconde après le début du nouveau fichier, j'ai également placé ce même bip de synchronisation une seconde après la fin de l'extrait, suivi de vingt-quatre images de silence. Le premier bip permettra de resynchroniser les fichiers, le deuxième permettra de s'assurer que le décalage observé est constant tout au long du fichier.

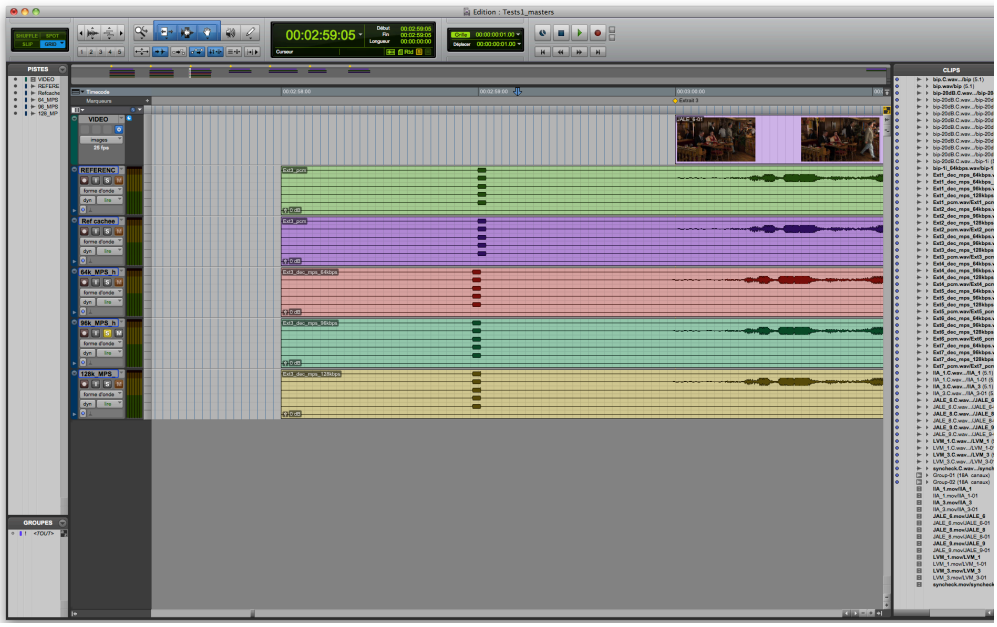


Figure 61 : Capture d'écran de ProTools : décalage des fichiers encodés

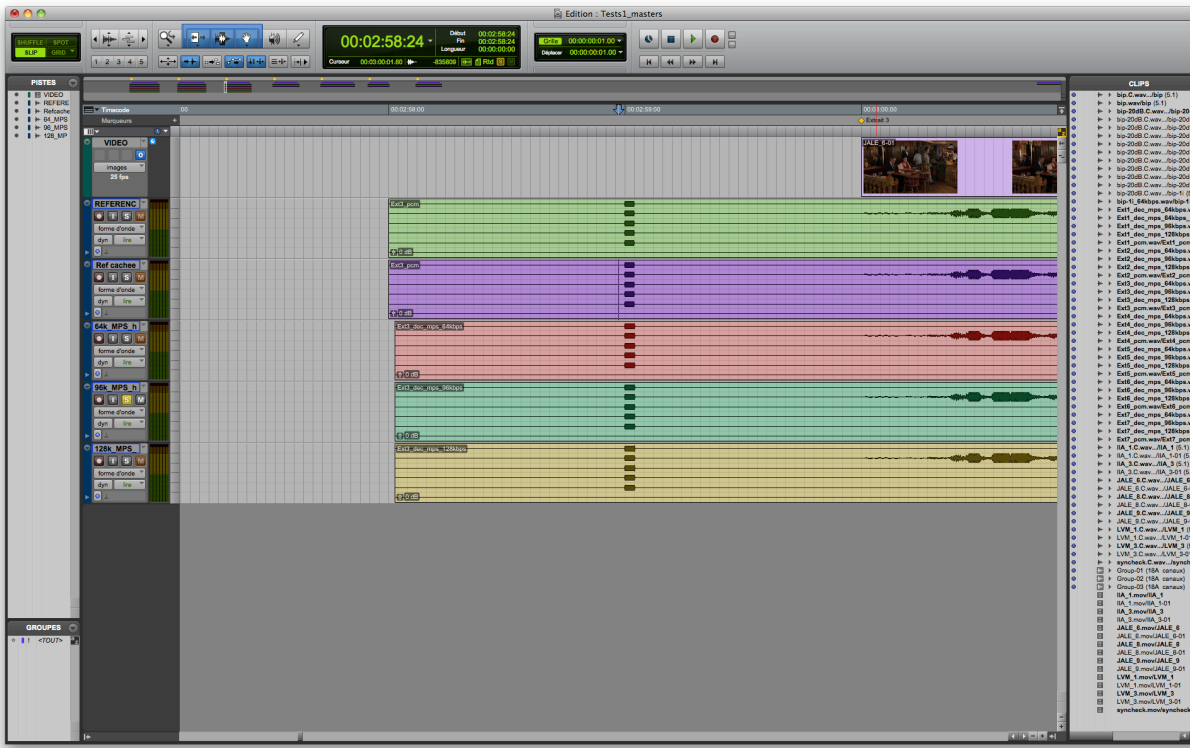


Figure 63 : Capture d'écran de ProTools : fichiers encodés recalés

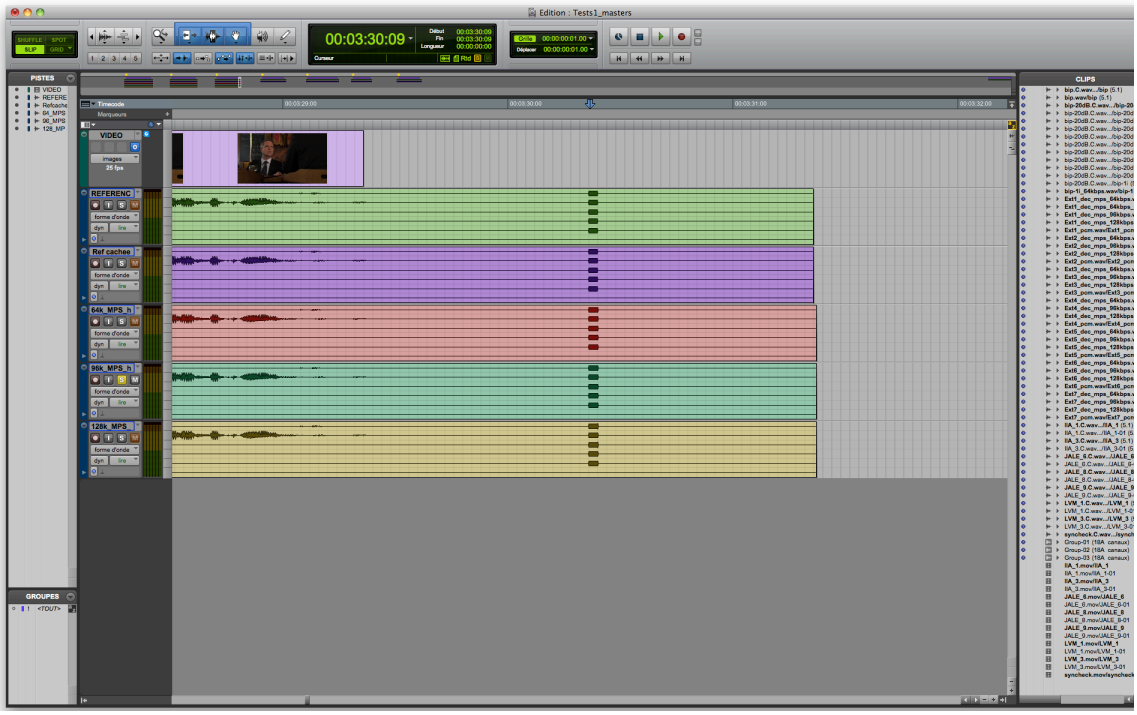


Figure 62 : Capture d'écran de ProTools : quand les 1ers bips sont synchrones, les bips de fin le sont aussi.

On observe qu'une fois que le premier bip est synchrone, le bip de fin l'est aussi : cela signifie donc que le début du fichier est tronqué (entre quelques millisecondes, jusqu'à 40 millisecondes, soit environ une image, sur des extraits de trente ou quarante secondes), mais une fois le premier bip synchronisé, le fichier est synchrone jusqu'à la fin, où on observe alors que des échantillons ont été ajoutés. Il me suffit donc d'encoder les fichiers avec deux bips de synchronisation, afin de les recaler après décodage, pour que le participant puisse switcher d'un stimulus à l'autre avec une continuité du son.

J'ai contacté la société Fraunhofer à propos de ce défaut, ils espèrent résoudre ce problème d'implémentation pour une prochaine version.

Le plugin propose en outre un encodage en binaural, néanmoins ce mode n'est pas très performant pour l'instant. Il s'agit d'un encodage particulier, différent d'un encodage traditionnel, il utilise une seule HRTF qu'on ne connaît pas et à laquelle on ne peut accéder, et propose un choix entre trois simulations de pièces (pièce « sèche », salon ou salle de cinéma). L'idéal serait évidemment de pouvoir intégrer nos propres HRTF. Il est donc préférable pour l'instant d'utiliser un autre encodeur binaural si on veut effectuer des tests plus poussés.

2. Lieu des Essais et Système d'Écoute

Les tests perceptifs ont eu lieu au laboratoire de France Télévisions, récemment construit en respectant toutes les préconisations de la recommandation ITU-R BS.1116-1 (issues des documents EBU-UER Tech. 3276 : *Listening conditions for the assessment of sound programme material* et document AES TD1001 *Multichannel surround sound systems and operations*).

2.1. Caractéristiques Géométriques et Acoustiques du Laboratoire

Le laboratoire mesure 5 mètres de large, par 6 mètres de long, et il a une hauteur de 2,7 mètres, soit une surface de 30 m^2 et un volume de 81 m^3.

L'écran mesure 3 mètres de largeur, il est constitué d'une toile acoustiquement neutre, les haut-parleurs frontaux et les subwoofers sont disposés derrière la toile. Le sweet-spot se situe à 2,50 mètres de l'enceinte centrale. Les enceintes gauche et droite sont espacées de 2,70 mètres, elles se situent à 30 cm du mur entre le nodal et le laboratoire (cf. figure 68). Les enceintes arrières sont positionnées à 60 cm des murs latéraux et à 1,20 mètre du mur arrière. Le vidéoprojecteur JVC, de la gamme D-ILA, projette une image en 2K.

Avant le traitement acoustique du local, voici les temps de réverbération globaux mesurés sur l'ensemble du spectre :

- ♪ Early Decay Time = 0,2 s
- ♪ T10 = 0,27 s
- ♪ T20 = 0,32 s
- ♪ T30= 0,47 s

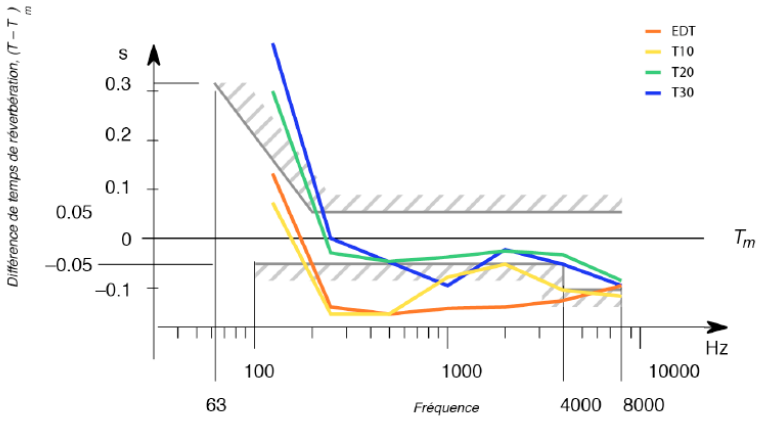


Figure 64 : Gabarit du temps de réverbération par octave, issu de la recommandation ITU-R BS.1116 – figure 1)

Tableau 3 : Conformité du laboratoire aux recommandations ITU

Caractéristiques	Laboratoire Le Ponant	Recommandation ITU-R BS. 1116 (pour système multi-voies)	Conformité
Longueur L (m)	6 m		/
Largeur l (m)	5 m		/
Hauteur h (m)	2,7 m		/
Surface (m^2)	6*5 = 30 m^2	Entre 30 et 70 m^2	ok
Volume (m^3)	6*5*2,7 = 81 m^3		/
Forme du local	Symétrique par rapport au plan vertical médiateur de la base stéréo. Surface au sol rectangulaire	Symétrique par rapport au plan vertical médiateur de la base stéréo. Surface au sol rectangulaire ou en trapèze.	ok
l/h	5/2,7 = 1,85	< 3	ok
L /h	5/2,7 = 2,22	< 3	/
1,1 l/h	1,1*5/2,7 = 2,04		/
4,5 l/h - 4	4,5*5/2,7 – 4 = 4,33		/
1,1 l/h < L/H < 4,5 l/h - 4	2,04 < 2,22 < 4,33 ?		ok
Tps moyen de réverbération entre 200 et 4 kHz	Tm = 0,25*(V/100)$^{1/3}$ Tm = 0,233 s	Tm à +/- 0,05 s	Ok pour le T20.
Bruit de fond ambiant	Supérieur à la courbe NR-20	Inférieur à la courbe NR-15	Non conforme
B (largeur de la base stéréo)	2,70 m	Entre 2 et 4 mètres	ok
D : distance d'écoute	2,50 m	D = B	≈
Angle au sweetspot	60°	60°	ok

Le temps moyen de réverbération Tm de la figure 64, est défini par la formule :

$$T_m = 0,25.\left(\frac{V}{V_0}\right)^{\frac{1}{3}}$$

où V est le volume de la pièce considérée et V_0 un volume de référence égal à 100 m^3. Ici le temps moyen de réverbération vaut 0,23 seconde.

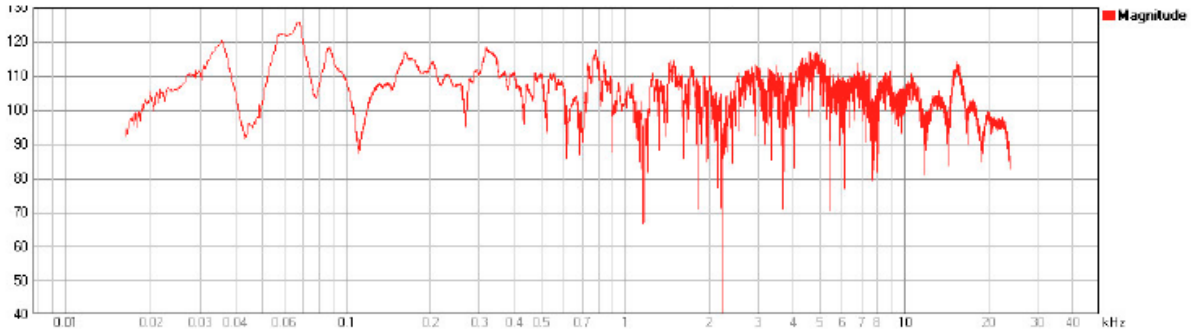


Figure 65 : Courbe de réponse en fréquence du laboratoire avant traitement

Les modes de cette pièce sont peu marqués, le mode présent à 63 Hz sera amorti. Des diffuseurs latéraux ont été placés pour améliorer la sensation d'espace.

L'isolation acoustique a été mesurée entre le laboratoire, le nodal, le bureau et le couloir. L'isolement acoustique standardisé et pondéré A est de 36,7 dB avec le bureau, de 33,7 dB avec la porte et de 35 dB avec le nodal. Ces valeurs d'isolation sont typiques de locaux administratifs, mais sont insuffisantes pour ce type de laboratoire, néanmoins l'activité alentour est relativement peu nuisible.

Le niveau de bruit résiduel du laboratoire est légèrement supérieur à la courbe de niveau résiduel NR-20, alors que les recommandations préconisent un niveau équivalent à la courbe NR-10, et ne devra excéder en aucun cas la courbe NR-15 : le laboratoire est donc trop bruyant. Néanmoins, les nuisances sont plutôt ponctuelles, comme par exemple le bruit de l'ascenseur ou le passage de personnes dans le couloir.

Quelques photos du laboratoire se trouvent sur la page suivante.

Figure 66 : Photos du laboratoire Le Ponant

Sur les pages suivantes figurent un schéma théorique de disposition du système d'écoute selon les recommandations ITU, ainsi qu'un plan du laboratoire.

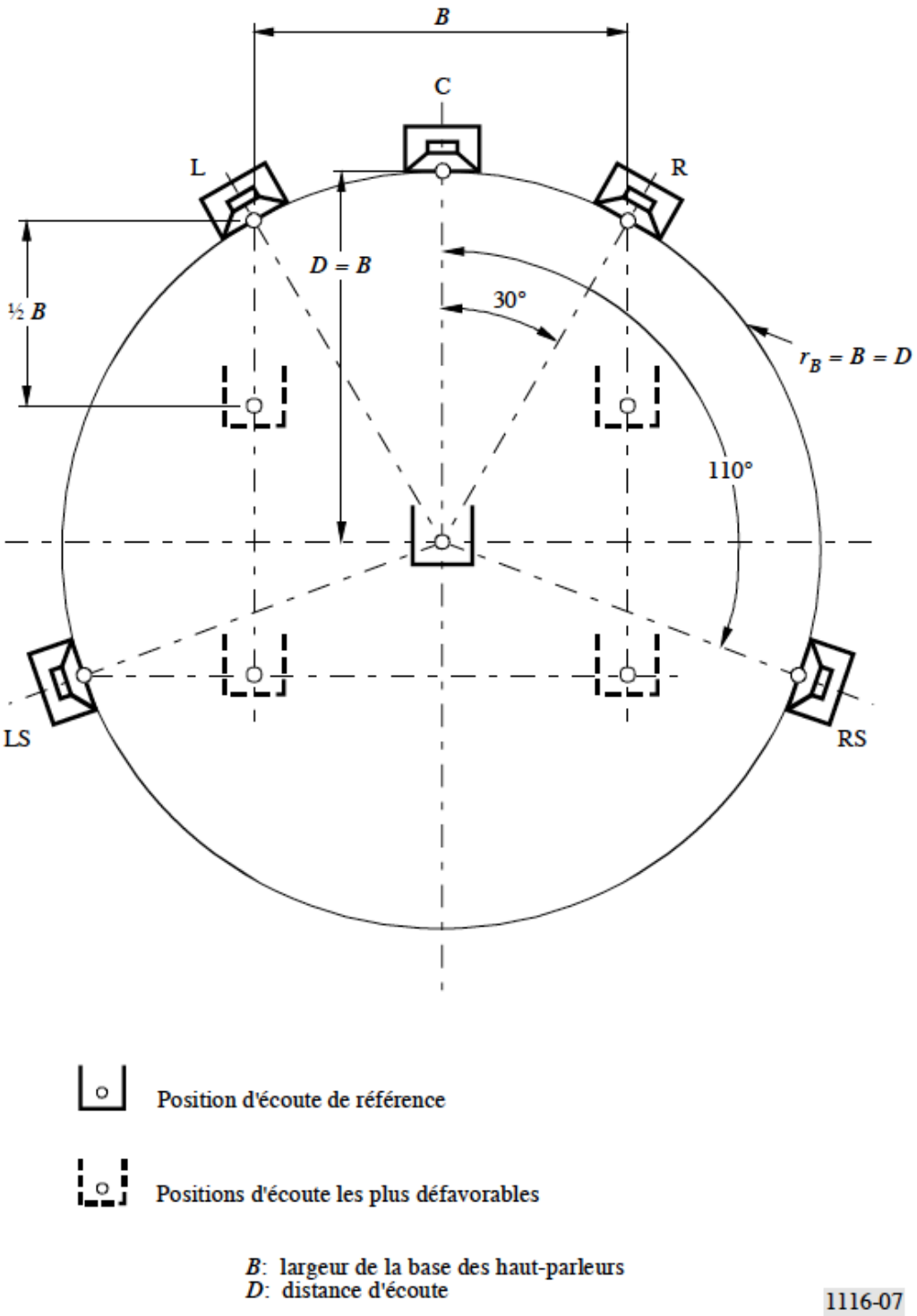


Figure 67 : Disposition d'écoute d'un système multivoies 3/2 selon la recommandation ITU-R BS.1116

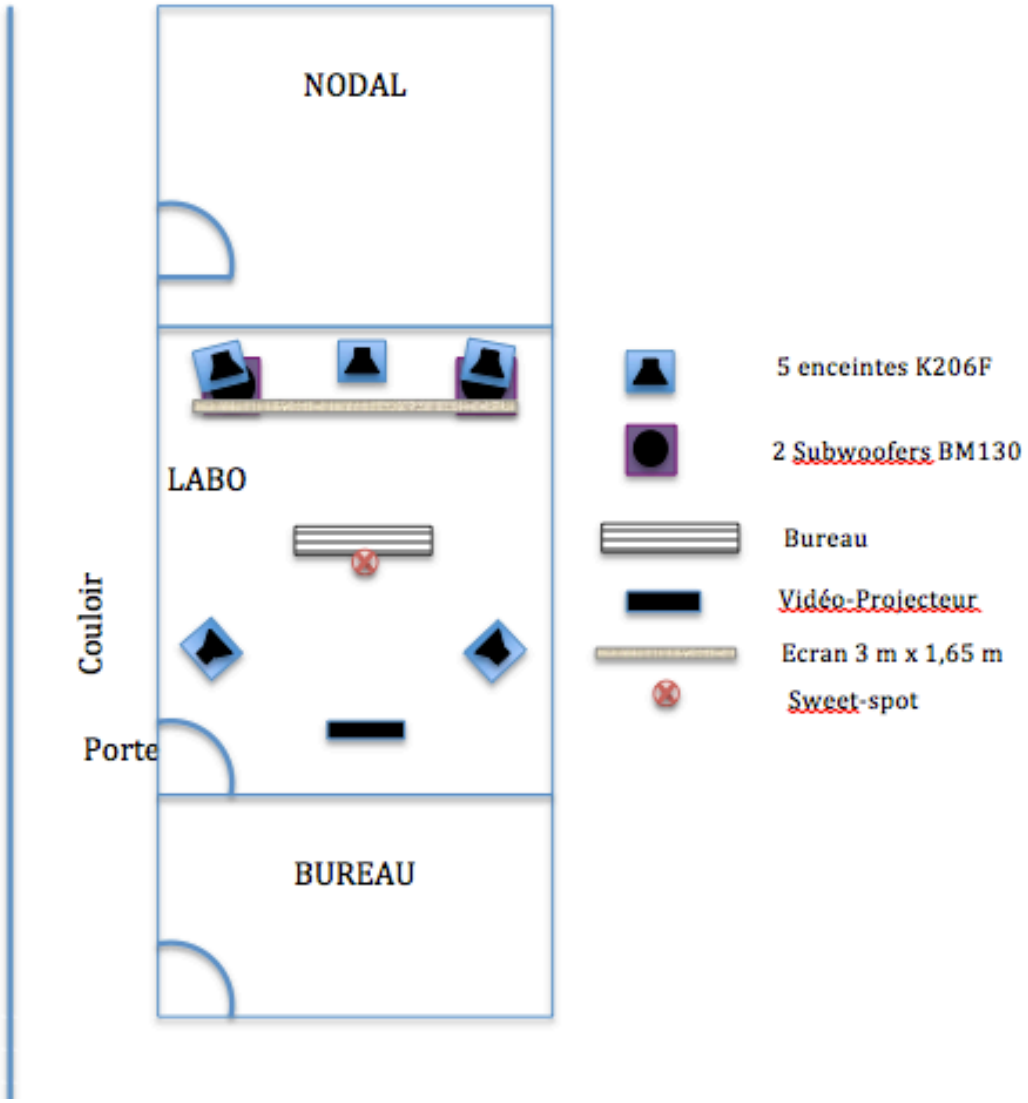


Figure 68 : Plan du laboratoire Le Ponant

2.2. Système d'Écoute

Le laboratoire dispose de plusieurs configurations sonores, on peut écouter de la stéréo, du son multicanal 5.1 et 7.1 et à terme il est destiné à proposer une écoute 22.2.

Les enceintes ont été construites de façon artisanale et sur mesure par Jérémie Guillemaut de D.J.E. Production. L'écoute en 5.1 (qui respecte la disposition préconisée

par l'ITU) se fait grâce à cinq enceintes K206F, et deux subwoofers BM130 (réglés de façon à ce que l'énergie sonore du canal LFE soit répartie entre les deux subwoofers).

Les enceintes K206F comporte deux haut-parleurs de 5 pouces Fostex filtrés passivement, et un tweeter à dôme. Les deux subwoofers BM 130 contiennent chacun un haut-parleur 12 pouces Beyma en charge Bass Reflex, ils ont une impédance de 8 ohms, et une réponse en fréquence de 25 à 800 Hz à +/- 4 dB, et une puissance RMS de 100 W.

Figure 69 : À gauche Subwoofer BM130, Au centre Enceinte K206F, À droite Enceinte Monitor Pocket

L'écoute en 22.2 se fait avec le système 5.1, auquel on ajoute seize enceintes Monitor Pocket, avec les mêmes haut-parleurs Fostex en large bande. Les enceintes Monitor Pocket ont les caractéristiques suivantes : un haut-parleur 5 pouces large bande Fostex en charge Bass Reflex, avec une réponse en fréquence de 60 Hz à 19 kHz, à +/- 4 dB, une impédance de 9 ohms, et une puissance RMS de 20 Watts et un niveau de 93 dB/1W/1m. Ainsi les vingt et une enceintes ont un timbre proche. Le choix d'utiliser des haut-parleurs large bande a été réalisé pour permettre une mise en réseau pour de la WFS plus tard, et pour limiter le poids de chaque enceinte sur la structure. (Il manque encore une enceinte arrière centre pour former un véritable 22.2).

Les amplificateurs se situent dans le nodal, il y a trois amplificateurs RAMaudio T1208 (huit voies chacun) et un amplificateur RAMaudio S3000 à deux voies pour les subwoofers.

2.3. MATÉRIEL

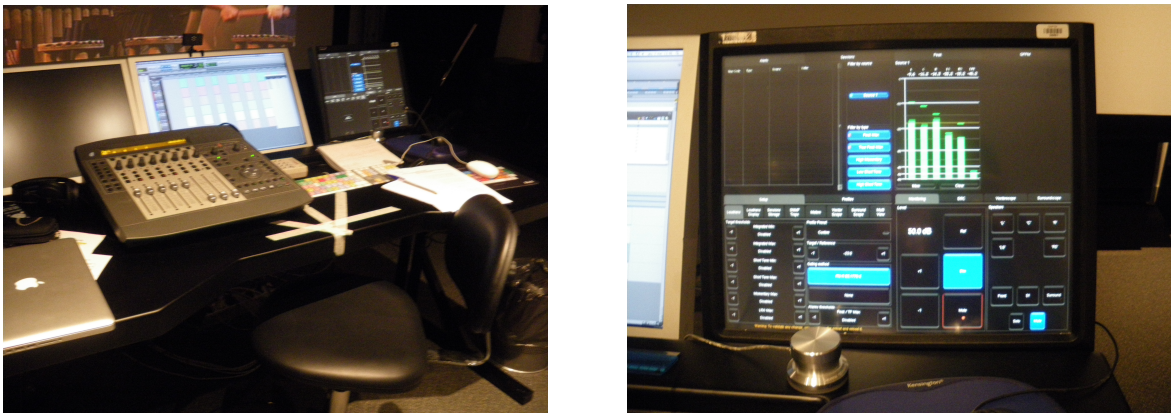


Figure 70 : Command 8 et interface du processeur d'écoute Trinnov

Je dispose d'un ProTools HD Native 10.3, relié à une interface MADI 64 canaux, ainsi que d'une surface de contrôle Command 8, qui permet au participant de gérer le transport, d'avancer dans la time-line et d'écouter la piste qu'il souhaite. Sur l'écran s'affiche le nom de la piste et le time-code au curseur. Le participant peut zoomer dans la time-line, peut réaliser des boucles plus courtes, mais il n'a pas le droit de voir ni le nom des clips, ni la forme d'onde.

Le processeur d'écoute Trinnov possède 24 canaux, il permet une adaptation électro-acoustique de l'écoute au local, ainsi que le réglage de la configuration et du niveau d'écoute, et peut aussi mesurer des niveaux sonores (notamment en mode loudness selon la recommandation R128). Ce processeur peut aussi simuler un récepteur de type grand public.

3. Première Série de Tests Perceptifs : Évaluer le Meilleur Débit d'Encodage en MPEG Surround

3.1. Mise en place du protocole et choix des méthodes

Le MPEG Surround peut introduire des dégradations faibles à intermédiaires de l'audio.

La recommandation ITU-R BS.1116-1 [33] décrit une méthode d'évaluation subjective de dégradations faibles de l'audio, méthode du doublement aveugle à triple stimulus et référence dissimulée. Pour chaque essai, le participant doit écouter trois stimuli : la référence connue, puis le signal dégradé et la référence cachée. La référence connue est toujours présentée et nommée stimulus « A ». La référence dissimulée et le signal dégradé sont alors présentés de façon aléatoire comme stimulus « B » ou « C. Le participant doit alors évaluer la dégradation du stimulus « B » comparé au stimulus « A » et du stimulus « C » comparé au stimulus « A » selon l'échelle continue de notation. L'un des deux stimulus « B » ou « C » ne devrait pas se distinguer du stimulus « A ». Toute différence perçue entre le stimulus de référence et le stimulus évalué est alors interprétée comme étant une dégradation. Cette méthode est particulièrement adaptée aux faibles dégradations audio.

La recommandation ITU-R BS.1534 [34] décrit quant à elle une méthode d'évaluation subjective de dégradations intermédiaires de l'audio, appelée méthode MUSHRA (Multi Stimulus test with Hidden Reference and Anchor) ou méthode « multi-

[33] ITU-R BS.1116-1 : *Méthodes d'évaluation subjective des dégradations faibles dans les systèmes audio, y compris les systèmes sonores multi-voies*, 1997.

[34] ITU-R BS.1534 : *Méthode d'évaluation subjective du niveau de qualité intermédiaire des systèmes de codage*, 2001.

stimuli avec référence et repère cachés », méthode ayant montré des résultats fiables pour évaluer la qualité audio intermédiaire de différents codecs et déjà utilisée pour évaluer la qualité du MPEG Surround. Pour chaque essai, le participant écoute la référence connue, ainsi que plusieurs signaux dégradés, la référence cachée et un ou plusieurs repères cachés : cette méthode permet de présenter tous les stimuli en même temps et l'auditeur peut alors faire les comparaisons de son choix. La durée du test est donc réduite par rapport à un test mené en double aveugle à triple stimulus. En effet, pour noter trois combinaisons en double aveugle à triple stimulus, on écoute donc trois fois trois signaux, soit neuf signaux. Si le même test est réalisé avec la méthode MUSHRA, le participant écoute alors la référence connue, puis les trois signaux dégradés et la référence cachée, voire un repère, ce qui fait au total cinq ou six signaux à écouter seulement. De plus, le participant doit pouvoir réaliser les comparaisons qu'il souhaite en gérant le transport. La durée totale du test varie donc en fonction du participant. Afin que le participant ne se fatigue pas trop, je limiterai la durée du test pour chaque extrait.

3.1.2. Son et/ou Image ?

Les différentes recommandations préconisent de réaliser des tests en son seul, afin que l'image ne perturbe pas et n'influence pas la notation du participant. Néanmoins, puisque mon sujet traite de l'application du MPEG Surround à une diffusion télévisuelle, il apparaît donc nécessaire de tester ce codage avec du son associé à l'image, afin d'évaluer outre le timbre et la qualité audio, l'adéquation entre la spatialisation du son et l'image. Cependant, nous avions convenu, que si lors de la première série de tests, on s'apercevait que les résultats obtenus n'étaient pas évidents, il faudrait alors envisager de reconduire ces mêmes tests sans image, afin de déterminer si l'image est un facteur discriminant.

Les méthodes de tests avec image d'accompagnement sont détaillées dans la recommandation ITU-R BS.1286[35], tandis que la recommandation ITU-R BS.1284[36] décrit diverses méthodes d'évaluation de la qualité audio.

3.1.3. Durée des Extraits et de la Séance

Selon la recommandation ITU-R BS .1116-1, une séance de notation ne devrait pas durer plus de vingt à trente minutes. Néanmoins, si le rythme de la séance est laissé à la discrétion du participant, le test pourra durer plus ou moins longtemps.

De plus, chaque extrait doit être homogène, c'est-à-dire par exemple une scène calme ou une scène d'action. En revanche, si au sein du même extrait, on a plusieurs coups de feu, puis une ambiance calme, on ne saura pas si l'auditeur a plutôt noté la reproduction de la dynamique des coups de feu, ou si au contraire il s'est plutôt fié aux ambiances calmes. Afin d'obtenir des avis fiables pour ces deux genres de scènes, il vaut mieux scinder l'extrait en deux parties.

De plus, la recommandation ITU-R BS.1534 préconise d'utiliser des extraits d'une durée inférieure à vingt secondes, car la mémoire auditive humaine est une mémoire à court-terme, et qu'il faut éviter de fatiguer les auditeurs. Cette durée est vraiment courte, et l'essentiel était de privilégier des mini-séquences qui se tiennent seules, pour ne pas couper au milieu d'une réplique. J'ai donc affiné le découpage de mes extraits sous ProTools, et je me suis donc aperçue que mes extraits avaient des durées d'environ trente secondes, à plus ou moins dix secondes. Avant de les encoder en MPEG Surround, j'ai dû ajouter à chaque extrait un bip de synchronisation à 3 kHz (bip issu de la séquence syncheck) d'une durée d'une image sur toutes les pistes, exceptée la piste LFE, bip placé une seconde après le début du fichier et une seconde avant le début de l'audio, et un autre

[35] ITU-R BS.1286 : *Méthodes d'évaluation subjective des systèmes audio avec image d'accompagnement*, 2007.

[36] ITU-R BS.1284 : *Méthodes générales d'évaluation subjective de la qualité du son*, 2003.

bip identique une seconde après la fin de l'audio et une seconde avant la fin du fichier. Puis j'ai exporté ces fichiers en 5.1 en mode entrelacé, je les ai ensuite ré-importés proprement dans la session ProTools. Ces bips de synchronisation me permettent de resynchroniser les fichiers décodés avec le fichier original.

Lors de ces premiers tests, je souhaite comparer trois débits différents encodés en HE-AAC (High-Efficiency Advanced Audio Coding). Ainsi, pour chaque extrait, le participant devra écouter le fichier original nommé Référence, puis quatre stimuli, parmi lesquels il y aura un signal original caché (référence cachée) et trois stimuli encodés en MPEG Surround, puis décodés, ce qui fait un total de cinq stimuli par extrait. En comptant une durée d'environ trente secondes, il faut alors deux minutes trente pour écouter les cinq stimuli en entier. Le participant sera libre de zapper d'un stimulus à l'autre, et de réécouter ceux qu'il souhaite, mais je limite la durée du test à cinq minutes par extrait. J'ai donc choisi d'utiliser sept extraits pour ce test, ce qui conduit à une durée maximale du test de trente-cinq minutes, certains iront plus vite, d'autres prendront ce temps-là.

3.1.4. Sélection des Participants

La recommandation ITU-R BS.1116-1 préconise de recourir à des auditeurs experts, qui ont l'habitude de détecter ce genre de défauts. Quand les conditions de l'expérience sont maîtrisées, une vingtaine de participants experts suffit pour obtenir de bons résultats. Néanmoins, en fonction des résultats, je pourrais être amenée à réaliser une post-sélection des participants si certains d'entre eux donnent des résultats incohérents par rapport à la majorité des participants, s'ils s'avèrent incapables de déceler la référence cachée à chaque essai, ou si certains sont jugés trop ou pas assez critiques. Le rejet d'un candidat dans les résultats sera explicité.

Puisque le temps est limité, j'ai donc lancé une première série de tests perceptifs avec une vingtaine de participants « experts », qui sont des camarades de l'école en

section son, quelques enseignants, ainsi que des assistants son et des ingénieurs du son de France Télévisions.

3.2. Description de la Méthode d'Essai

Pour ces tests, je me suis inspirée de la méthodologie MUSHRA, décrite dans la recommandation ITU-R BS.1534.

Le participant disposera d'un signal original en format WAVE, nommé explicitement « référence », et de quatre signaux nommés « A », « B », « C », et « D » à noter, parmi lesquels il y aura une référence cachée, dont la note devrait être égale à 10, ainsi que trois signaux encodés de façon différente, puis décoder, et pour lesquels le participant doit apprécier un critère. Contrairement à la méthodologie MUSHRA, il n'y aura pas d'ancre cachée : une ancre est en fait un signal original, auquel on a appliqué un filtrage passe-bas à 3,5 kHz, en suivant la recommandation. L'ancre permet alors à l'auditeur d'avoir un repère de notation. Néanmoins, recourir à une ancre ajoute un signal de plus à évaluer, le test durera plus longtemps, et la bande-passante préconisée de 3,5 kHz apparaît faible, et sans grand intérêt à comparer avec le même signal original encodé en MPEG Surround, même au plus faible débit. On a donc choisi de ne pas utiliser d'ancre dans ces premiers tests.

3.2.1 Choix des Paramètres d'Encodage

Pour chaque extrait, je propose trois combinaisons débit-format identiques, à savoir MPEG-Surround combiné au codeur fondamental HE-AAC à 64, 96 et 128 kbits/seconde, afin de déterminer, pour chaque extrait, le débit minimal à utiliser, et ainsi déduire si on pourrait trouver un débit convenable pour tout type de programmes, ou si au contraire certains types de programmes ou de séquences sont beaucoup plus exigeants que d'autres en terme de débit.

Bien que le plugin Fraunhofer Pro-Codec de Sonnox propose un encodage en temps réel ainsi qu'un mode de comparaison en aveugle, mais inadapté à mes tests, il est préférable, pour des questions de ressources et de facilité de déroulement de l'expérience, d'encoder au préalable chaque extrait à ces différents débits en mode « offline », puis de les décoder, et de les importer dans la session.

Dans le mode « Offline encode » (cf. figure 71), on importe le fichier 5.1 original entrelacé, puis on règle les paramètres d'encodage : le codec choisi est MPEG Surround combiné au codeur HE-AAC, le débit est de 64, 96 ou 128 kbits/seconde, en mode débit constant (CBR ; il existe aussi un mode VBR ou débit variable, incompatible avec ce codec), en haute qualité (donc encodage lent), le fichier aura bien une fréquence d'échantillonnage de 48 kHz comme l'original, et enfin la sortie du décodeur sera dans le même format que le fichier original, donc ici un fichier 5.1 entrelacé. On aurait pu choisir une qualité encore meilleure (« highest quality ») mais le fichier encodé aurait eu une fréquence d'échantillonnage de 44,1 kHz, et aurait donc du être reconverti lors de l'importation du fichier décodé. La figure 72 représente les paramètres d'encodage choisis en fonction des débits.

Figure 71 : Capture d'écran du plugin Fraunhofer Pro-Codec - Mode Offline Encode

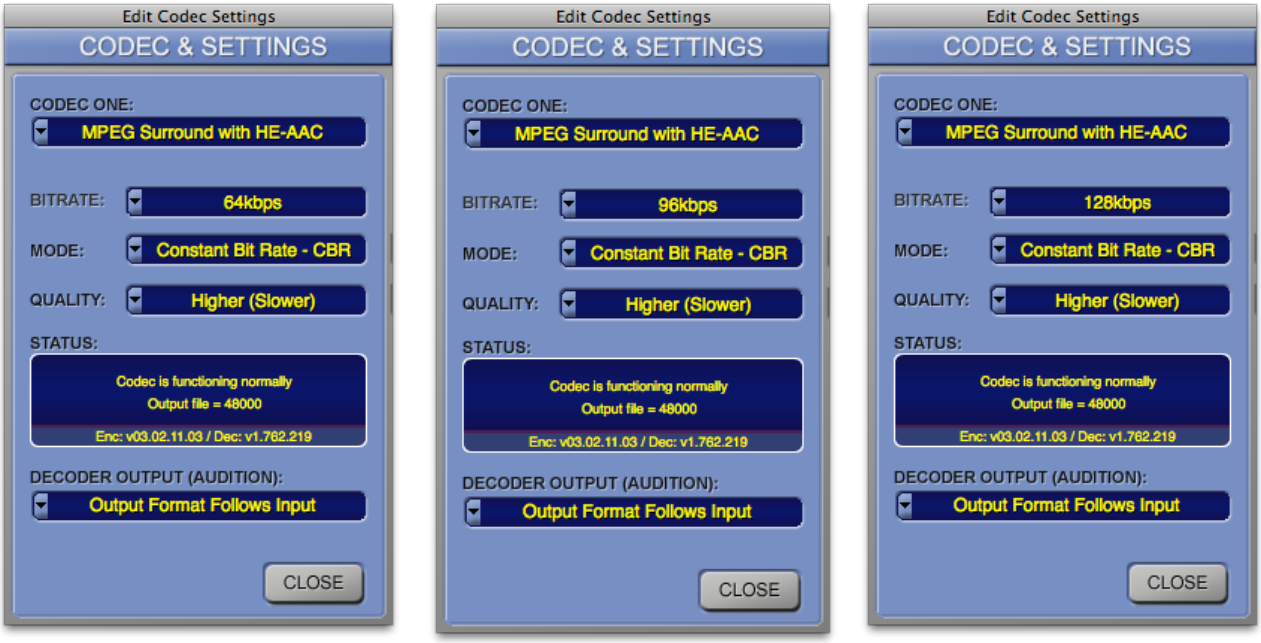


Figure 72 : plugin Fraunhofer ProCodec : paramètres d'encodage

Après encodage, on obtient des fichiers ayant pour extension .m4a, qui sont des flux MPEG-4 contenant le downmix audio stéréophonique et un flux de données de spatialisation, d'un débit compris entre 10 et 50 kbits/seconde, qui dépend du débit total et du contenu du fichier original.

J'ai ensuite décodé tous les fichiers. Pour cela, j'ai importé les flux MPEG-4 les uns après les autres, j'ai choisi de les décoder en format WAVE en 24 bits, afin de conserver les mêmes caractéristiques que les fichiers originaux (cf. figure 73). J'obtiens alors des fichiers entrelacés en 5.1. Ces fichiers sont ensuite importés dans ProTools et représentent donc les fichiers encodés et décodés à un certain débit en MPEG Surround combiné à HE-AAC. Les fichiers étant tous en 5.1 entrelacés, je peux alors les placer sur la time-line, les re-synchroniser avec le fichier original (afin de pallier l'artefact de décalage temporel apporté par l'implémentation du codec dans ce plug-in), et re-découper tous les fichiers, originaux et décodés, afin de supprimer les bips de synchronisation, et que tous les stimuli fassent la même durée.

Figure 73 : Capture d'écran du plugin Fraunhofer Pro-Codec - Mode Offline Decode

3.2.1. Phase de Familiarisation

La recommandation ITU-R BS.1534 conseille de réaliser une phase de familiarisation avec les participants, afin qu'ils intègrent la méthode de notation, ainsi que les types de dégradations qu'ils vont rencontrer, sur les mêmes extraits qu'ils devront ensuite noter : cette phase d'entraînement prend du temps et néglige l'aspect du ressenti et de la première impression, puisque le participant connaît alors déjà certains extraits. Il a été décidé de ne pas recourir à cette phase de familiarisation, afin de ne pas rallonger le test. Les éventuelles erreurs de notation au début du test, qui pourraient être dues au temps d'adaptation à la méthodologie d'évaluation, seront diluées grâce à un ordre aléatoire de présentation des extraits et des stimuli.

3.2.3. Choix des Extraits et Critères de Notation

J'ai donc sélectionné sept extraits d'une durée d'environ trente secondes. Parmi tous les extraits exportés, j'ai donc conservé deux extraits de l'opéra *L'Italiana in Algeri*, trois extraits du téléfilm *Jusqu'à l'enfer* de Denis Malleval, et deux extraits du documentaire *La vie Moderne* de Raymond Depardon. Les extraits durent entre vingt-trois secondes et quarante secondes.

Tableau 4 : Description des extraits du test 1

Extrait	Programme	Type	Durée	Description
1	*L'Italiana in Algeri* 1er acte	Opéra filmé	33s 21i	Ouverture de l'opéra. A l'image: fosse d'orchestre, chef d'orchestre, gros plans sur certains instruments. Au son: orchestre frontal, quelques nuances, des percussions (triangle, cymbales, etc.). Réverb' étendue sur les arrières.
2	*L'Italiana in Algeri* 1er Acte	Opéra filmé	32s 09i	Fin du duo entre Lindoro (ténor) et Mustafa (basse) sur l'air « Se inclinassi a prender moglie », puis applaudissements. A l'image: scène avec les 2 chanteurs, l'un tape sur des genres de tabourets. A la fin, les deux chanteurs s'en vont, et plan large du décor. Au son: orchestre + voix sur les enceintes avant (gauche/centre/droite), réverb sur les arrières. Applaudissements répartis principalement sur les canaux arrières
3	*Jusqu'à l'enfer* de Denis Malleval	Téléfilm	29s 09i	Scène au restaurant, à midi, dialogue entre 3 acteurs. Ambiance de restaurant (post-synchronisée) centrée. Canaux arrières très faibles.
4	*Jusqu'à l'enfer* de Denis Malleval	Téléfilm	29s 16i	Scène en pleine campagne de nuit. Acteur seul qui appelle à l'aide. Ambiance nocturne calme, brise légère, en bordure de route. Passage de plusieurs voitures, dans différentes directions. Quelques basses dans le LFE au passage de voitures.
5	*Jusqu'à l'enfer* de Denis Malleval	Téléfilm	39s 22i	Scène de fin. Séquence de nuit. A l'image: alternance maquette du train et véritable train, le héros avance sur la voie de chemin de fer. Au son : musique (image stéréo large), sons de la maquette et du véritable train, sound design, pas sur les cailloux, klaxons du train. Canal LFE chargé.
6	*La vie moderne* de Raymond Depardon	Documentaire	23s 20i	Séquence de jour avec un berger et ses moutons. A l'image: la campagne, les moutons, le berger. Au son : ambiance de campagne, cloches des moutons, voix in du berger, et voix-off centrée. Peu d'arrières. Pas de LFE.
7	*La vie moderne* de Raymond Depardon	Documentaire	28s 22i	Séquence devant une ferme, de jour. A l'image : plan fixe sur la ferme, le paysan entre dans le cadre par la droite, marche jusqu'à la ferme, entre puis ferme la porte. Au son : on entend le paysan arriver par l'arrière droit, déplacement précis jusqu'à la porte d'entrée, qu'on entend s'ouvrir puis se fermer. Ambiance de campagne calme, au loin quelques voix et le son d'un briquet.

Pour chaque extrait, le participant devra donner une appréciation de la qualité audio globale, caractéristique unique qui tient compte de toutes les différences perçues entre le signal original et le signal à évalué. Les recommandations ITU-R BS.1116-1 et ITU-R BS.1534 préconisent de n'évaluer qu'un seul critère par séance, ceci peut apparaître réducteur, néanmoins si l'on pose plusieurs questions par extrait, comme par exemple la reproduction du timbre ou de la dynamique, la qualité frontale de l'image sonore (c'est-à-dire la qualité des sources sonores frontales et leur spatialisation, la perte de définition), la qualité d'impression ambiophonique (impression d'enveloppement) etc., le risque est de déconcentrer le candidat, il va se perdre par rapport à tous les points qu'il doit évaluer, et finira par répondre au hasard, ce qui donnerait des résultats peu fiables. Il est donc préférable de ne poser qu'une seule et même question pour tous les extraits, et elle sera la suivante : donner une appréciation globale de l'audio. La question est large mais permet d'obtenir de véritables informations quant au ressenti global du participant : d'ailleurs, c'est bien ce qui est primordial pour ce test. Si je veux affiner l'expérience, je poserai d'autres questions lors de prochains tests.

Pour la notation, la méthodologie MUSHRA préconise d'utiliser l'échelle continue ci-contre. Pour noter les extraits sur le critère de l'appréciation globale de l'audio, les participants disposeront donc de cette échelle graduée de 0 à 10, avec la présence des termes décrivant les intervalles uniquement sur la première page (mauvais de 0 à 2, médiocre de 2 à 4, assez bon de 4 à 6, bon de 6 à 8 et excellent de 8 à 10). Pour chaque stimulus de chaque extrait, nommés « A », « B », « C » et « D », ils placeront une croix sur cette droite, qui déterminera la note du stimulus. La référence cachée, puisqu'elle est identique au signal original, doit donc obtenir la meilleure note, soit 10. Si le participant trouve deux signaux excellents et proches de l'original, il peut leur attribuer la note de 10 à tous les deux : cela signifiera que le signal encodé apparaît de qualité égale au signal original.

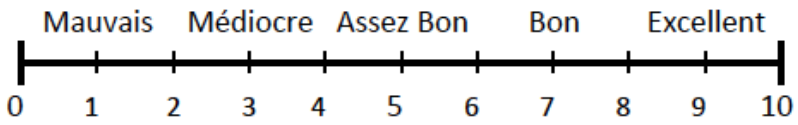


Figure 74 : Échelle de notation utilisée pour le test1

La figure 74 illustre l'échelle graduée utilisée pour la notation de ce test : elle est régulièrement graduée de 0 à 10, et mesure 10 cm de longueur : il me suffira alors de prendre une règle graduée pour déterminer la note attribuée à chaque stimulus, avec pour échelle : 10 cm équivaut à la note 10/10, donc 1 cm équivaut à la note 1/10 et 0,1 cm équivaut à la note de 0,1/10. Cette échelle est suffisamment large, et en même temps pas trop graduée, pour ne pas influencer les participants, tout en leur laissant suffisamment de marge pour différencier la notation des stimuli.

3.2.4. Réalisation de la Session Masters et des Sessions de Tests

Une fois les paramètres et les extraits choisis, je débute alors la session Masters sous ProTools 10. J'encode alors les fichiers originaux en MPEG Surround en mode Offline Encode, je conserve précieusement les flux MPEG-4, puis je les décode à l'aide de ce même plugin, en mode Offline Decode. Ma session Masters se présente ainsi : au time-code 00 :01 :00 :00 débute l'extrait 1 original, à savoir l'extrait d'opéra nommé IIA_1, au time-code 00 :02 :00 :00 débute l'extrait 2 original, à savoir l'extrait d'opéra nommé IIA_3, au time-code 00 :03 :00 :00 débute l'extrait 3 original, à savoir l'extrait du téléfilm nommé JALE_6, et ainsi de suite.

Ma session Masters, session de départ, est composée d'une piste vidéo et de cinq pistes 5.1, la première piste est appelée « Référence » et contient tous les signaux de référence de chaque extrait, la deuxième piste est une copie conforme de la première piste et contient donc les stimuli nommés Référence cachée. La troisième piste contient tous les stimuli décodés issus d'un encodage en MPEG Surround + HE-AAC à 64 kbits/seconde, la quatrième piste quant à elle contient tous les stimuli décodés issus d'un encodage en MPEG Surround + HE-AAC à 96 kbits/seconde, et enfin la cinquième piste contient tous les stimuli décodés issus d'un encodage en MPEG Surround + HE-AAC à 128 kbits/seconde. La figure 75 est une capture d'écran de la session Masters.

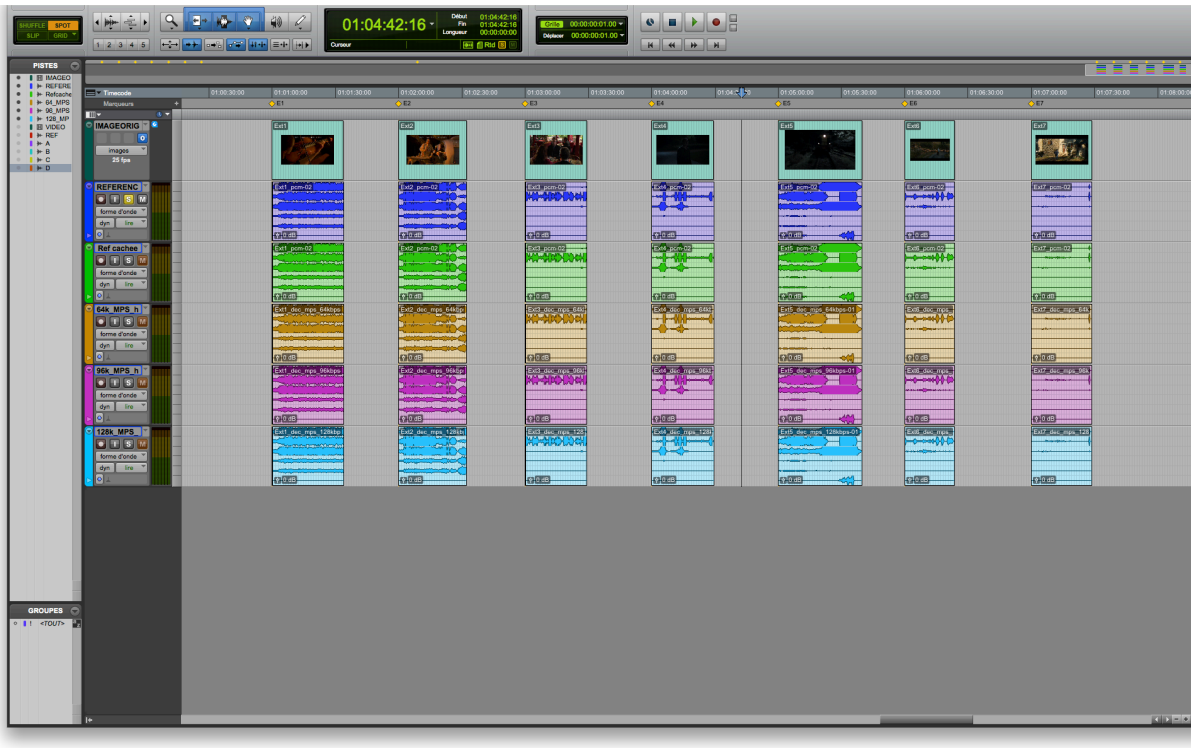


Figure 75 : Session Master - ProTools

Une fois la session Masters prête, je peux alors générer une quarantaine de sessions (au cas où j'aurais plus de participants que prévu), afin que les extraits et les stimuli soient présentés d'une façon aléatoire d'un participant à l'autre. Ceci présente plusieurs avantages : d'abord, cela permet d'éviter le recours à une phase d'entrainement, et dilue les erreurs dues à la découverte du mode de notation. Puisque les extraits sont présentés dans un ordre aléatoire, cela limite aussi les impacts des erreurs dues à la fatigue des auditeurs sur la fin du test, puisque le dernier extrait à noter n'est jamais le même. La présentation aléatoire des stimuli limite les habitudes que pourrait prendre un participant qui aurait remarqué que la piste 2 renferme toujours la référence cachée, chaque stimulus n'est donc jamais écouté avant ni après le même encodage.

La répartition des extraits et des stimuli de façon aléatoire a été définie par un programme de génération de séquences aléatoires, il me suffira de noter en haut du questionnaire le numéro de la session utilisée, et je pourrai directement remplir un tableau de résultats. Toutes les sessions tests se présentent de la même façon : une piste vidéo, une piste 5.1 nommée clairement « Référence » et quatre autres pistes 5.1 nommées

« A », « B », « C », et « D », qui contiennent les trois stimuli décodés et la référence cachée, répartis de façon aléatoire entre ces pistes. Les formes d'ondes ainsi que les noms des clips sont masqués.

La figure 77 montre une capture d'écran d'une session de tests quelconque, prête à l'emploi. Les noms des clips, le « clip gain », et les formes d'ondes ont été masqués. Chaque piste est correctement nommée. Des marqueurs définissent le numéro de l'extrait, afin que le candidat puisse se repérer, et la fenêtre des emplacements mémoires est ouverte, afin de faciliter le passage d'un extrait à l'autre.

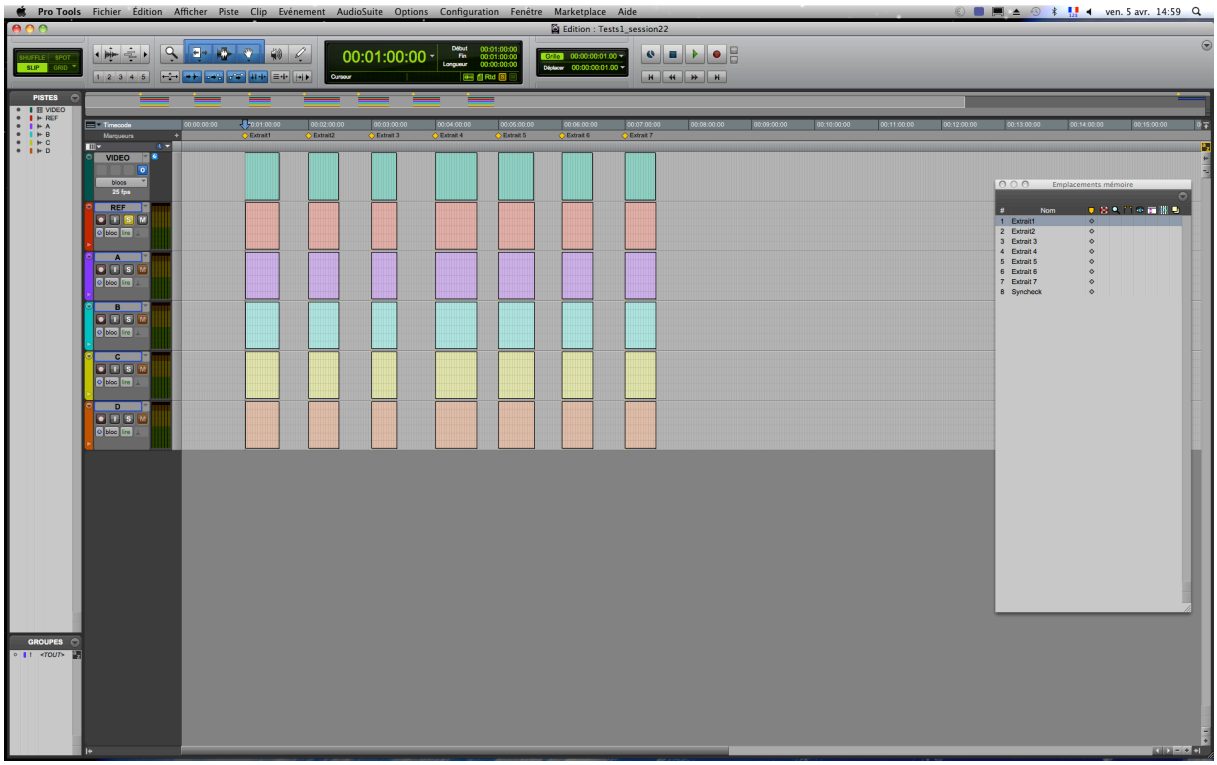


Figure 76 : Session X prête - Test 1

3.2.5. Déroulement de la Séance

Le participant a donc accès à l'écran du ProTools mais ne peut voir ni la forme d'onde, ni le nom des clips. Il est libre d'écouter en entier chaque stimulus, ou au contraire de zapper d'un stimulus à l'autre. L'interface Command 8 lui permet de gérer le transport. Le logiciel est paramétré afin que le solo soit en mode solo X-or : ainsi, dès que le participant active le solo sur une piste, le solo précédent est effacé, c'est-à-dire que la piste précédemment à l'écoute est automatiquement « mutée », avec un petit fondu. Pour zapper d'un stimulus à l'autre, il n'a donc besoin que des boutons Solo de la Command 8, et éventuellement de la souris d'ordinateur pour réécouter un passage précis. Sur l'interface Command 8, en haut de chaque fader, figure le nom de la piste, pour les sessions tests, dans l'ordre : Référence, A, B, C, D ; ainsi que le time-code au curseur, très utile pour se repérer dans la session.

Le participant dispose de cinq minutes pour chaque extrait : dans ce temps imparti, il doit avoir écouté tous les stimuli au moins une fois, et les avoir notés. S'il termine avant, il peut directement enchainé avec l'extrait suivant. C'est l'expérimentateur qui chronomètre chaque extrait.

Les consignes du test soumises aux participants se trouvent en Annexe B, page 197.

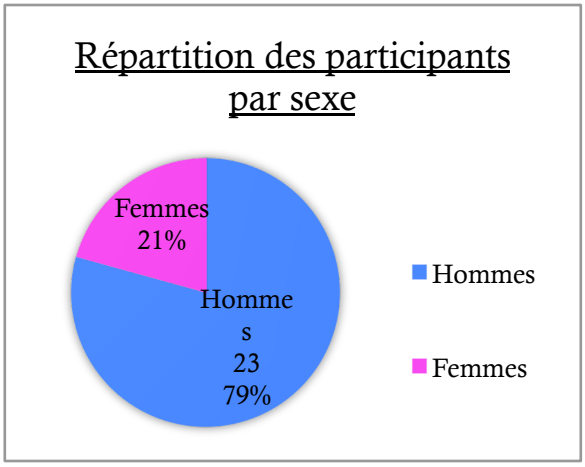
Le niveau sonore sera réglé à 70 dB sur l'appareil de monitoring Trinnov, et les candidats ne pourront pas le modifier. Ce niveau paraît un peu fort pour un des extraits, mais bien adapté à mon sens pour les autres extraits. Un niveau de 68 dB serait trop faible pour les autres extraits.

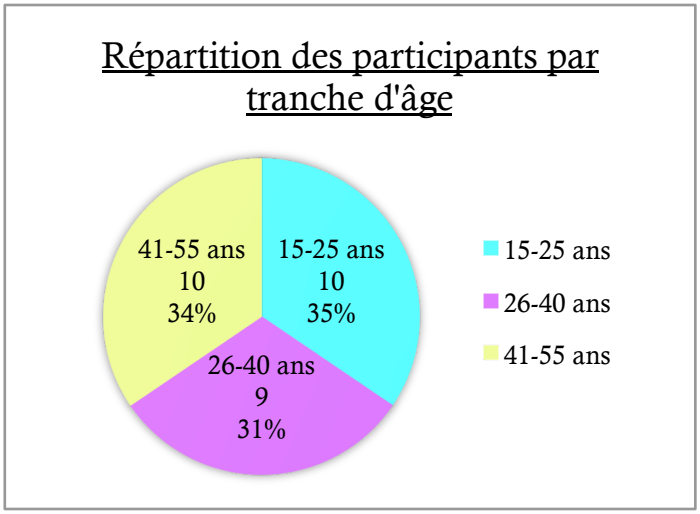
4. Analyse des Résultats des Premiers Tests

4.1 Participants

Au total, vingt-neuf participants ont participé à ce test, parmi eux six femmes et vingt-trois hommes, ayant entre vingt ans et cinquante-cinq ans.

Parmi eux, on dénombre neuf étudiants en formation son, huit ingénieurs du son pour la télévision (opérateurs de prise de son, chefs opérateur du son, mixeurs), sept personnes travaillant dans les domaines de l'ingénierie ou responsables audio de différents services, deux personnes ayant eu une expérience professionnelle dans l'audio, ainsi que trois personnes extérieures au monde de l'audio, mais musiciens amateurs. Neuf participants déclarent avoir l'habitude des tests d'écoute.





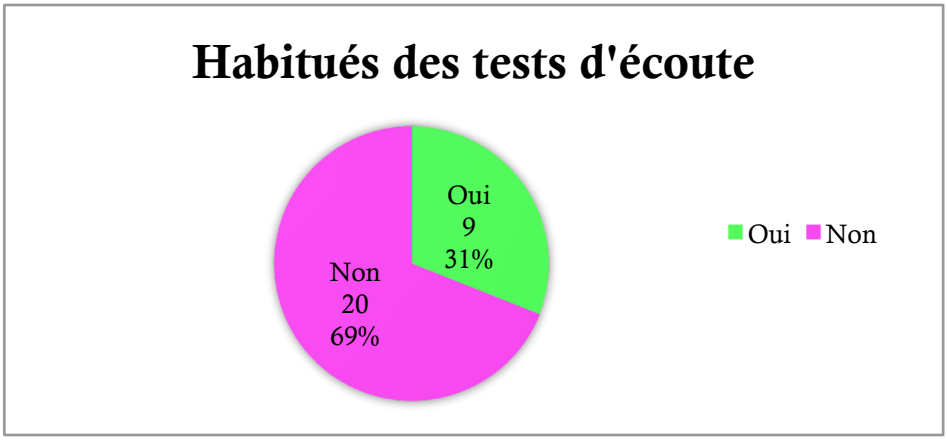


Figure 77 : Caractéristiques des participants

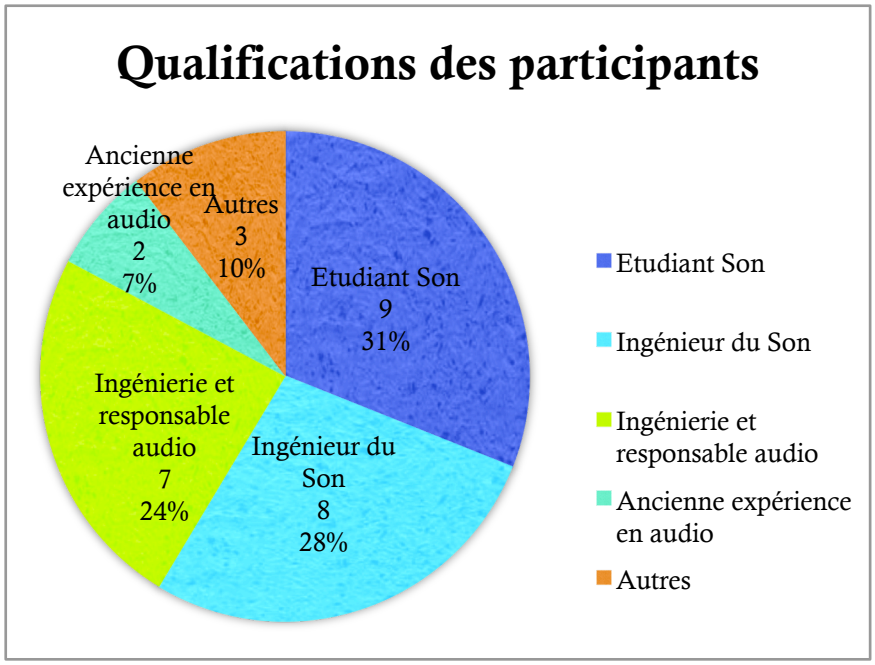


Figure 77bis : Qualification des participants

Huit participants possèdent un système home-cinéma chez eux pour écouter la télévision, six écoutent la télévision seulement avec les haut-parleurs de la télévision, huit autres l'écoutent avec deux enceintes externes et/ou casque, et certains utilisent plusieurs systèmes audio. Huit participants ne détiennent pas de téléviseurs, mais certains d'entre eux regardent tout de même quelques programmes télévisés en direct ou en replay depuis leur ordinateur.

4.2 Analyse des Résultats

4.2.1. Résultats Globaux Avant Post-sélection

Parmi les vingt-neuf participants, un des participants, considéré comme « non expert », n'entendait pas de différences entre les différents signaux et n'a donc pas donné de notes. Ce candidat est donc éliminé d'office.

Afin d'avoir un premier aperçu d'ensemble, j'ai calculé des premières valeurs. Toutes les notes données par les vingt-huit candidats se trouvent en annexe B, pages 198 à 204, classées par extrait et par codec.

Tableau 5 : Moyennes globales par codec, tous extraits confondus

Tous extraits confondus				
CODEC	Référence cachée	MPS 64 kbits/s	MPS 96 kbits/s	MPS 128 kbits/s
MOYENNE PAR CODEC	8,69	7,35	7,94	8,05
ECART-TYPE (σ_{N-1})	1,909	2,293	1,968	2,000
Note minimale	2,000	1,100	2,000	2,000
1er Quartile	8,000	6,000	7,000	7,000
Médiane	10,000	8,000	8,000	9,000
3ème quartile	10,000	9,000	9,750	10,000
Note maximale	10,000	10,000	10,000	10,000

Avant post-sélection, on remarque que le codec HE-AAC + MPEG Surround à 64 kbits/seconde obtient la moyenne la plus faible, de 7,35/10, mais l'écart-type est élevé. Les moyennes obtenues par les codecs HE-AAC + MPEG Surround à 96 kbits/seconde et 128 kbits/seconde sont très proches, avec un écart-type presque identique. La référence cachée obtient la meilleure moyenne, mais seulement égale à 8,69/10. Or, cette moyenne aurait dû être très proche de 10/10. On remarque d'ailleurs qu'au moins un candidat a mis une note de 2/10 à la référence cachée : ce candidat n'a donc pas été apte à reconnaître la référence cachée, et devra être éliminé. De plus, la moyenne de la référence cachée démontre la difficulté des candidats à discerner la référence cachée des autres

codecs : ceci explique en partie pourquoi les moyennes des différents codecs sont peu différenciées.

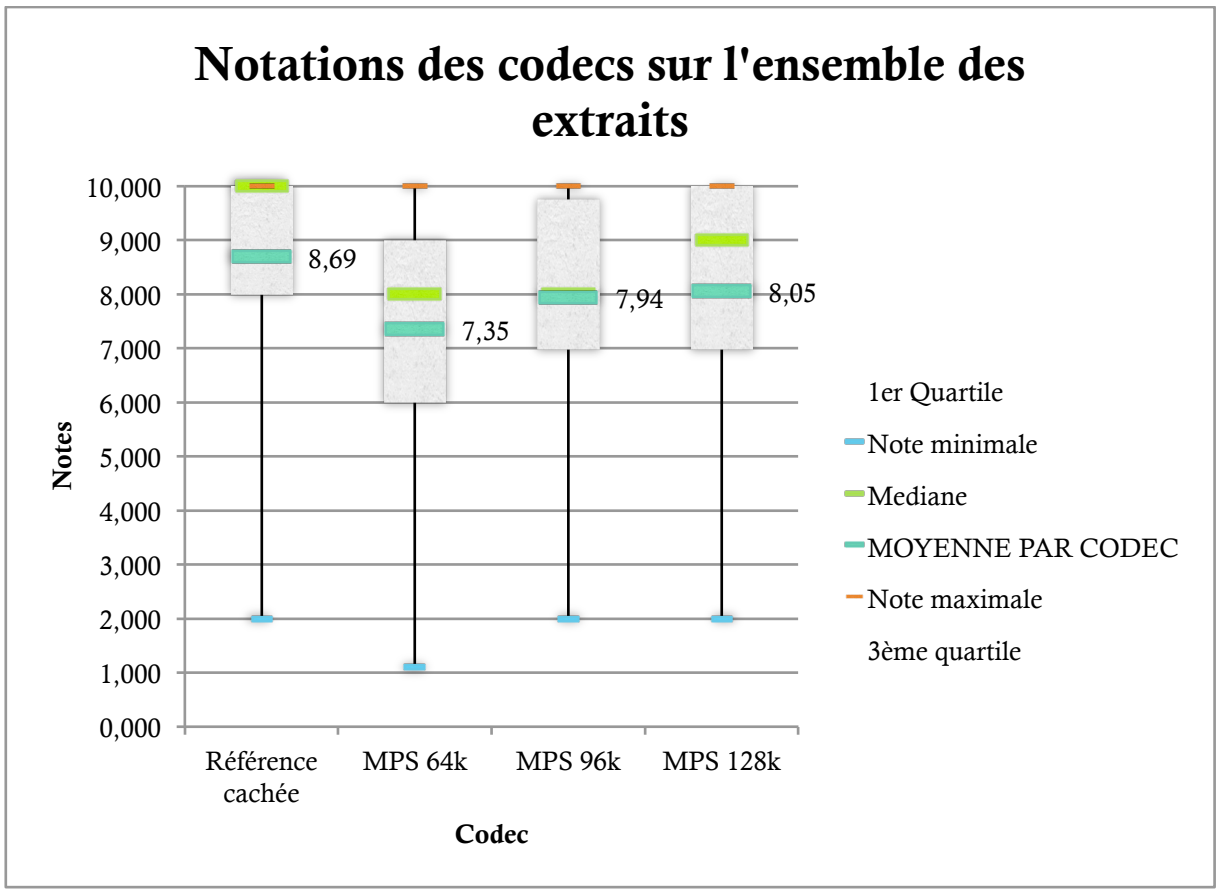


Figure 78 : Notations des codecs sur l'ensemble des extraits

L'extrait 1 est issu de l'ouverture de l'opéra *L'Italiana in Algeri*, la référence cachée obtient des notes comprises entre 5 et 10, avec une moyenne de 8,62. On peut remarque que les trois codecs ont obtenu au moins une note égale à 10. La codec au plus faible débit obtient la plus faible moyenne, mais l'écart-type est très étendu : les moyennes ne sont donc pas significativement différentes entre les trois codecs et la référence cachée. Ici, les principaux artefacts repérables sont surtout un timbre un peu moins riche dans les hautes fréquences, et une légère réduction de la largeur stéréophonique.

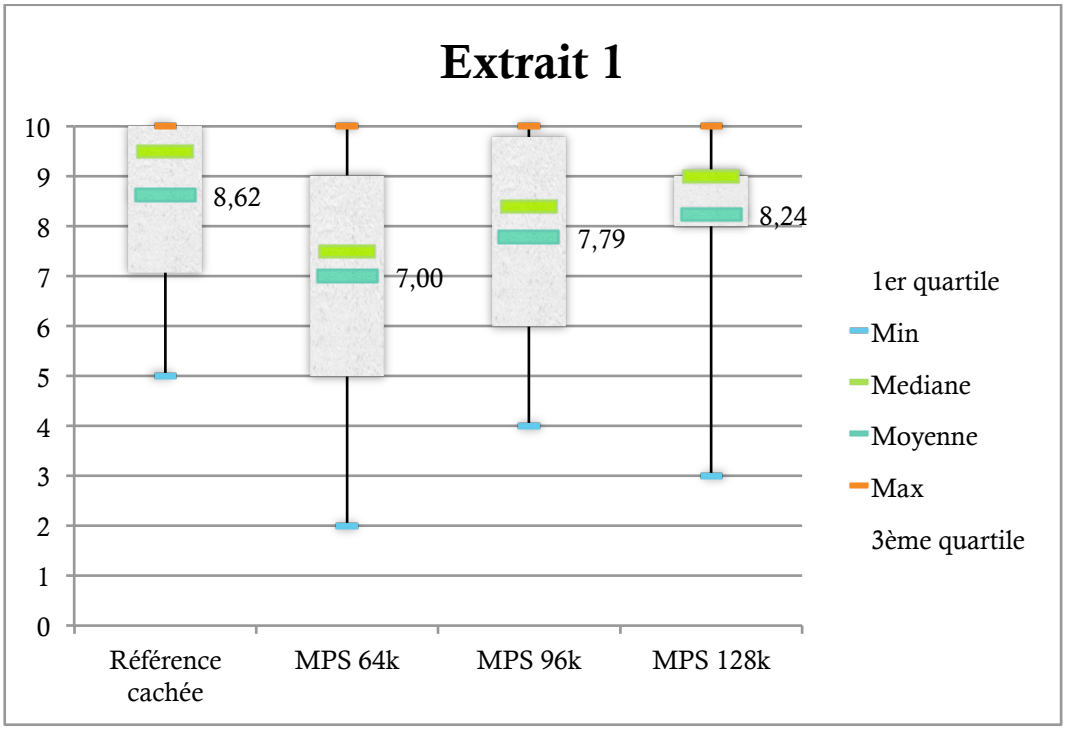


Figure 79 : Notations de l'extrait 1

L'extrait 2 issu de l'Acte 1 de l'opéra *L'Italiana in Algeri* contient la fin d'un duo, puis des applaudissements. Ici, on remarque que la référence cachée a une moyenne très proche de 10 : elle a donc été très bien détectée. Par contre, les trois codecs ont une moyenne très proche et ne sont donc pas différenciés, d'ailleurs c'est le codec à 96 kbits/seconde qui obtient la plus mauvaise moyenne. Les applaudissements sont

particulièrement critiques : le timbre est dégradé, et les applaudissements se déplacent entre l'avant et l'arrière, c'est probablement eux qui ont aidé la notation. Les timbres des instruments et des voix sont eux aussi dégradés.

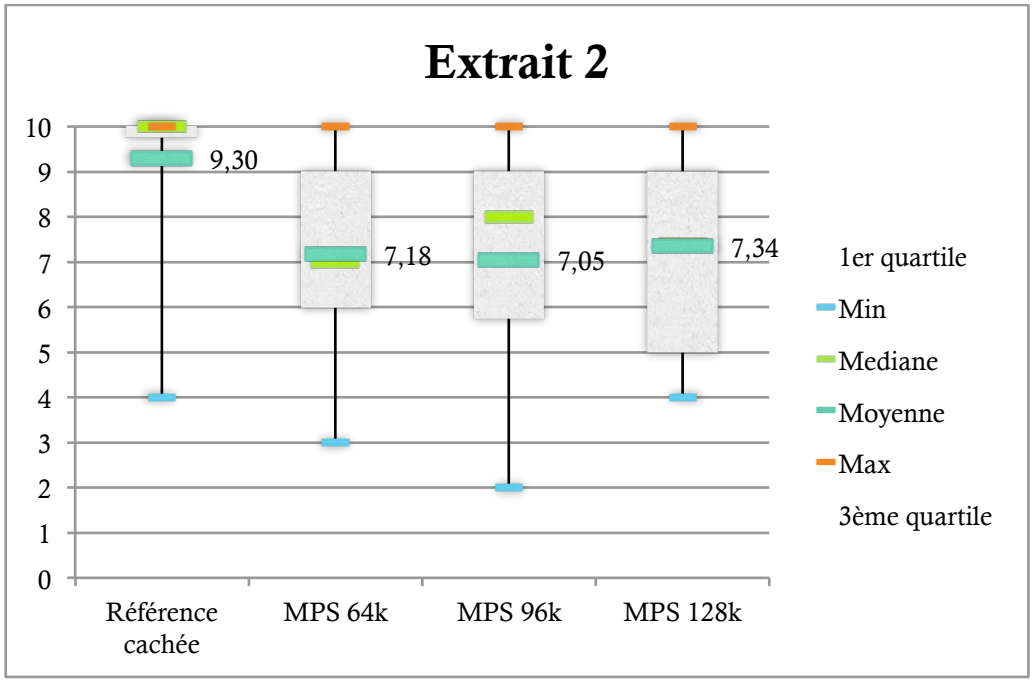


Figure 80 : Notations de l'extrait 2

Dans l'extrait 3, on assiste à une scène dans un restaurant, avec un dialogue et une ambiance post-synchronisée centrés, et très peu d'ambiances dans les canaux arrières. Cet extrait est particulièrement représentatif de la difficulté des candidats à différencier la référence cachée des trois codecs puisque deux des codecs obtiennent une meilleure moyenne que la référence cachée, qui obtient seulement un 8,55/10. De plus, on remarque que pour chaque signal, de nombreux candidats (plus de 25%) ont mis une note de 10/10. Il sera intéressant de comparer les moyennes après post-sélection.

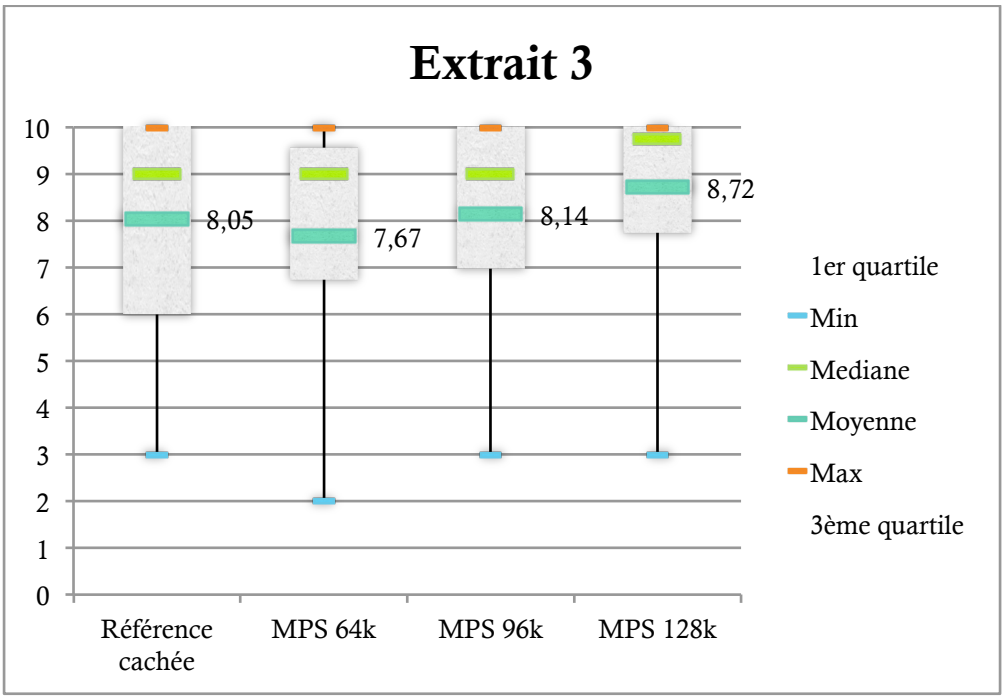


Figure 81 : Notations de l'extrait 3

L'extrait 4 est une séquence de nuit, en campagne, avec des passages de voiture. Tous les codecs ainsi que la référence cachée ont des moyennes très proches : les différences ne sont pas significatives. Le déplacement des voitures a été plutôt bien reproduit, quelques artefacts sont perceptibles dans les fonds d'air.

L'extrait 5 est l'extrait le plus long, issu de la séquence finale du téléfilm *Jusqu'à l'enfer*, qui mêle une maquette du train et le train réel, avec une musique très présente, ayant une image stéréophonique très large. La référence cachée obtient la meilleure moyenne : 8,91/10, et les trois codecs ont des moyennes très proches, peu différenciés, mais exprimant une « bonne qualité » (entre 6,75 et 7,56). Néanmoins, chaque codec obtient au moins une note très basse et une note égale à 10. Le codec 96 kbits/seconde obtient la meilleure moyenne. Cet extrait est l'un des extraits les plus discriminants car on sent une nette réduction de l'image stéréophonique sur la musique, et des artefacts sur les

effets spéciaux. Néanmoins, la réduction de l'image stéréophonique est certes détectable sur une écoute dont les enceintes gauche et droite sont espacées de 2,70 mètres, mais je ne suis pas sûre que la sensation soit autant évidente dans un petit salon.

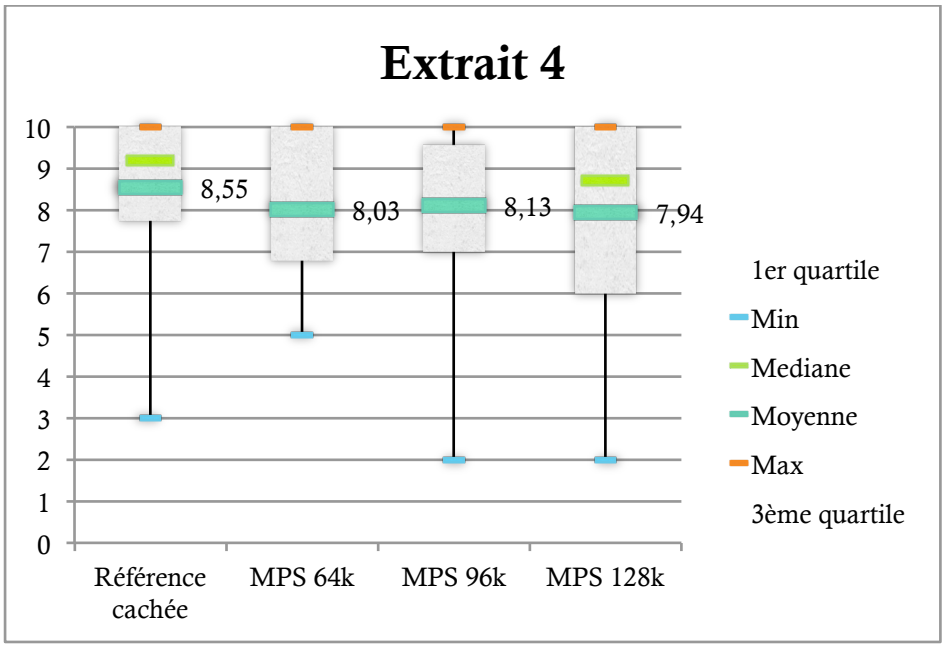


Figure 82 : Notations de l'extrait 4

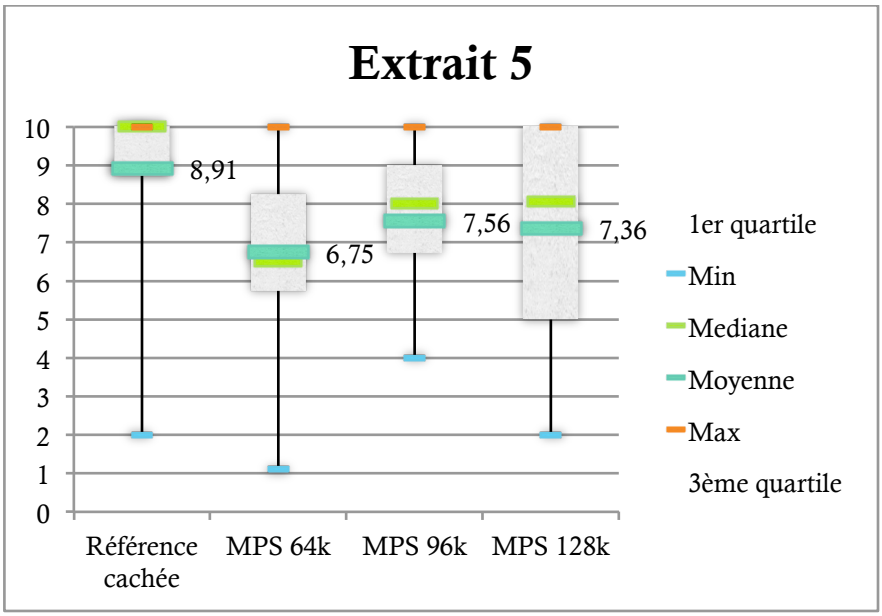


Figure 83 : Notations de l'extrait 5

L'extrait 6 est issu du documentaire *La vie moderne*, il s'agit d'une séquence avec des moutons qui se déplacent, on entend leurs cloches, et une voix off. Ici, on remarque que les moyennes de tous les codecs sont très rapprochées, d'ailleurs c'est le codec à 96 kbits/seconde qui obtient la meilleure moyenne. Cet extrait a été difficile à analyser, de très légères différences sur les cloches sont perceptibles par les oreilles les plus fines, notamment un timbre légèrement différent et une précision moindre pour les signaux encodés. A l'exception du codec à 64 kbits/seconde, plus de 50% des candidats ont mis une note supérieure ou égale à 9 à tous les autres codecs (y compris la référence cachée).

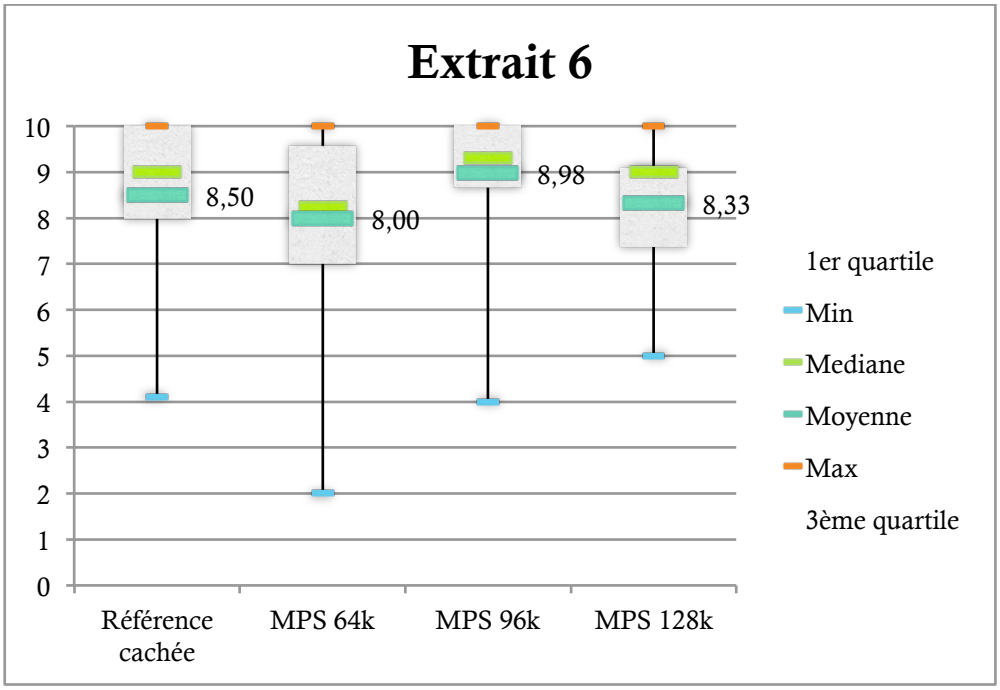


Figure 84 : Notations de l'extrait 6

Le dernier extrait comporte un déplacement précis d'un paysan, issu du documentaire *La vie moderne*. Cet extrait permettait d'évaluer précisément la reproduction de la spatialisation. La référence cachée obtient la meilleure moyenne : 8,93/10. Le codec à 64 kbits/seconde a la moyenne la plus faible : 6,86/10, tandis que le codec à 128 kbits/s

a la moyenne la plus élevée : 8,45/10. Le déplacement a été très bien reproduit, des différences infimes dans les fonds d'air sont perceptibles mais très peu les ont discernées.

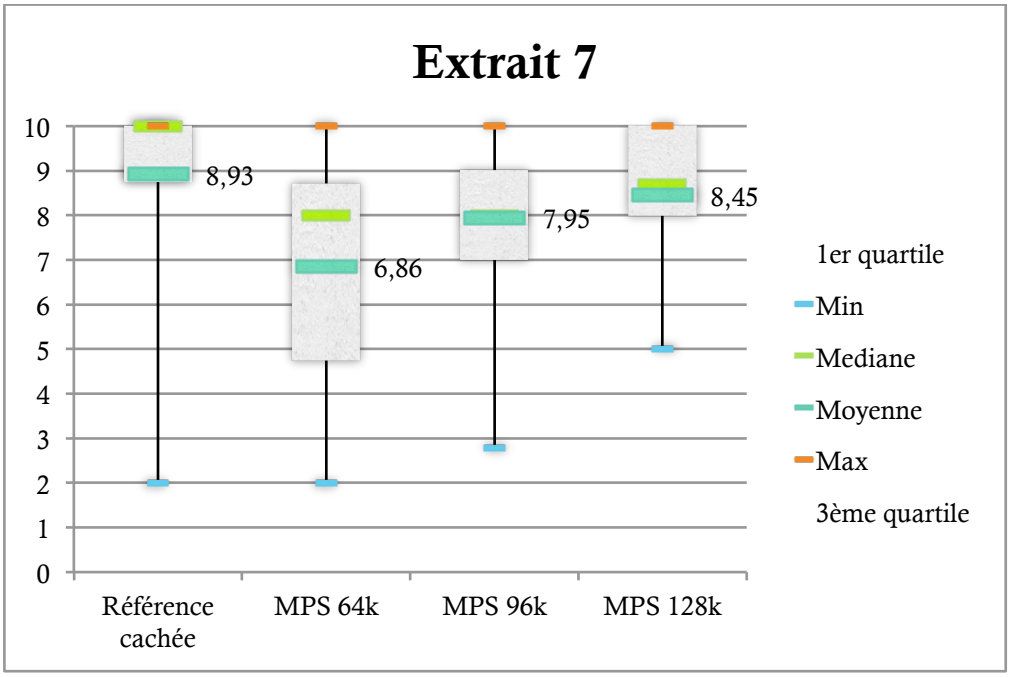


Figure 85 : Notation de l'extrait 7

4.2.2. ANOVA de Kruskal Wallis

Une ANOVA de Kruskal Wallis a été réalisée rapidement sur l'ensemble des codecs et des extraits, dont voici les résultats :

Hypothèse H0 : les échantillons proviennent d'une même population.

Hypothèse Ha : les échantillons proviennent de populations différentes.

Au seuil de 5%, puisque la probabilité calculée est inférieure à 0,05, on doit rejeter l'hypothèse H0 et retenir l'hypothèse Ha : les échantillons proviennent de population différentes. Ceci montre déjà une incohérence dans les résultats obtenus.

Tableau 6 : ANOVA de Kruskal Wallis

K (Valeur observée)	50,953
K (Valeur critique)	7,815
DDL	3
p-value (bilatérale)	< 0,0001
alpha	0,05

Tableau 7 : Comparaisons multiples par paires

	Référence cachée	MPS 64k	MPS 96k	MPS 128k
Référence cachée		**Oui**	**Oui**	**Oui**
MPS 64k	**Oui**		Non	**Oui**
MPS 96k	**Oui**	Non		Non
MPS 128k	**Oui**	**Oui**	Non	

Une comparaison multiples par paires suivant la procédure de Steel-Dwass-Critchlow-Fligner a permis de différencier trois groupes distincts (cf. tableau 7) : la référence cachée est donc bien distinguée des autres codecs. En revanche, le codec HE-AAC + MPEG Surround à 96 kbits/seconde est à la fois proche du codec HE-AAC + MPEG Surround à 64 kbits/seconde et du codec HE-AAC + MPEG Surround à 128 kbits/seconde alors que ces deux derniers sont différenciés : cela signifie donc que les trois codecs ont des moyennes très proches, et donc pas significativement différentes.

De plus, un test de Shapiro-Wilk a été effectué pour chaque codec, tous extraits confondus : il s'avère qu'aucune distribution ne suit une loi normale, alors qu'on peut considérer qu'une notation comprise entre 0 et 10 correspond à une variable continue. Mais je ne peux utiliser le théorème central limite[37] puisque je n'ai que vingt-huit participants. Il va donc être difficile de calculer de façon simple un intervalle de confiance.

[37] Le théorème central limite autorise, pour toute distribution d'une variable aléatoire dont le nombre d'échantillons (ou ici de candidats) est supérieur à 30, de faire l'approximation que la distribution suit une loi normale. Cette approximation facilite les calculs, notamment d'intervalles de confiance.

On peut tout de même noter que quelque soit les extraits, la référence cachée a obtenue plusieurs notes très basses et n'a donc pas été distinguée par certains candidats, tandis que tous les autres codecs ont obtenu plusieurs 10/10 : soit les candidats n'ont pas entendu de différences, soit ils ont trouvé le signal d'excellente qualité.

Ces résultats ne sont donc pas très exploitables en tant que tels, il faut procéder à une post-sélection pour affiner l'analyse.

4.2.2. Analyse des Erreurs sur la Référence Cachée et Élimination de Certains Candidats

Avant de commencer toute analyse statistique, il faut d'abord procéder à l'élimination des candidats les moins aguerris.

Ayant suivi la méthodologie MUSHRA, avec une référence cachée pour chaque extrait, que les candidats devaient noter à 10, on peut penser que les candidats ayant commis trop d'erreurs sur la référence cachée doivent être éliminés. Hélas, ce n'est pas si évident.

Par exemple, si on observe le nombre d'erreurs relatifs sur la référence cachée quand elle obtient une note inférieure à au moins un des autres codecs, on s'aperçoit que seulement douze candidats sur les vingt-huit, soit 43% des participants, ont commis au maximum deux erreurs sur la référence cachée, et ont donc un taux de réussite supérieur à 70% (cf. figure 86).

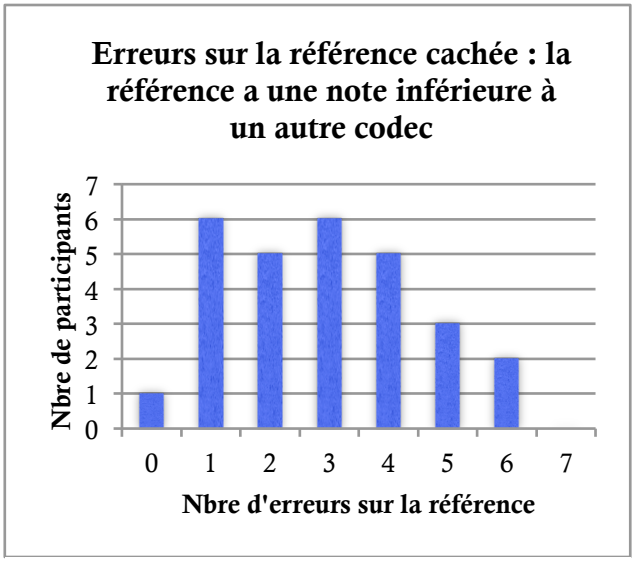


Figure 86 : Nombre d'erreurs sur la référence cachée : un autre codec obtient une meilleure note

En revanche, si on considère le nombre d'erreurs absolues sur la référence cachée, c'est-à-dire quand la référence cachée obtient une note strictement différente de 10, on

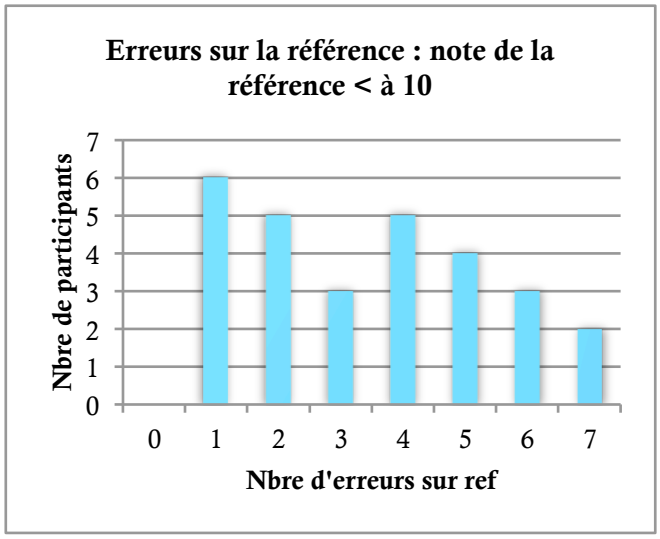


Figure 87 : Erreurs sur la référence cachée (elle obtient une note inférieure à 10)

remarque alors qu'il n'y a plus que onze des candidats qui ont un taux de réussite supérieure à 70%, c'est-à-dire qui ont commis au maximum deux erreurs.

Si on veut garder davantage de candidats, on augmente alors le nombre acceptable d'erreurs commises : mais trois erreurs sur sept extraits mène tout de même à un taux d'erreurs de 43%, c'est-à-dire presque une fois sur deux : à ce niveau d'erreurs-ci et sur aussi peu d'extraits, on peut quasiment considérer que le candidat a répondu au hasard.

La recommandation ITU-R BS.1116 préconise de calculer la différence entre la note d'un codec et la note de la référence cachée, et ce pour tous les extraits et pour chaque candidat. Si la moyenne de ces différences tend vers 0, cela signifie que le candidat a eu de grandes difficultés à discerner la référence cachée et a potentiellement mis des notes au hasard. Au contraire, si la moyenne des différences s'éloigne de zéro du côté négatif, cela signifie que le candidat a été dans l'ensemble, capable de discerner la référence cachée.

J'ai alors calculé les différences entre chaque codec et la référence cachée, et ce pour chaque extrait et chaque participant. J'ai calculé la moyenne et l'écart-type des différences pour chaque participant (cf. Annexe B pages 207 à 209). J'ai ensuite réalisé sous XLSTAT un test paramétrique bilatéral t de Student[38], afin de déterminer la probabilité que la différence des notes ait une moyenne égale à 0, au seuil de 10% et de 25%.

[38] On fait l'approximation que les différences de notation suivent une loi normale.

Le test t de Student révèle donc que quatorze participants ont une probabilité inférieure à 10% d'obtenir une moyenne des différences égales à 0, et dix-sept participants ont une probabilité inférieure à 25% d'obtenir une moyenne des différences égale à 0 : cela signifie que ces dix-sept participants, au seuil de confiance de 25%, n'ont pas essayé de deviner quel signal était la référence cachée, mais ont réellement entendu des différences, qu'ils ont évaluer.

Je retiens donc ce critère, et les participants que je garde pour l'analyse suivante sont les dix-sept participants ayant une probabilité inférieure à 25% d'obtenir une moyenne des différences égale à 0.

4.2.3. Évaluation de la Qualité des Codecs

Après post-sélection, je retiens dix-sept candidats, et j'ai donc recalculé les moyennes et écart-types.

Tableau 8 : Notations des codecs après la 1ère post-sélection

Notations codecs (tous extraits confondus) après la 1ère post-sélection				
CODEC	Référence cachée	MPS 64kbits/s	MPS 96 kbits/s	MPS 128 kbits/s
MOYENNE PAR CODEC	9,12	7,27	7,63	8,09
ECART-TYPE (σ_{N-1})	1,540	2,088	1,868	1,941
Note minimale	3,000	2,000	2,800	2,000
1er Quartile	9,000	6,000	6,000	7,000
Médiane	10,000	7,500	8,000	9,000
3ème quartile	10,000	9,000	9,000	9,750
Note maximale	10,000	10,000	10,000	10,000

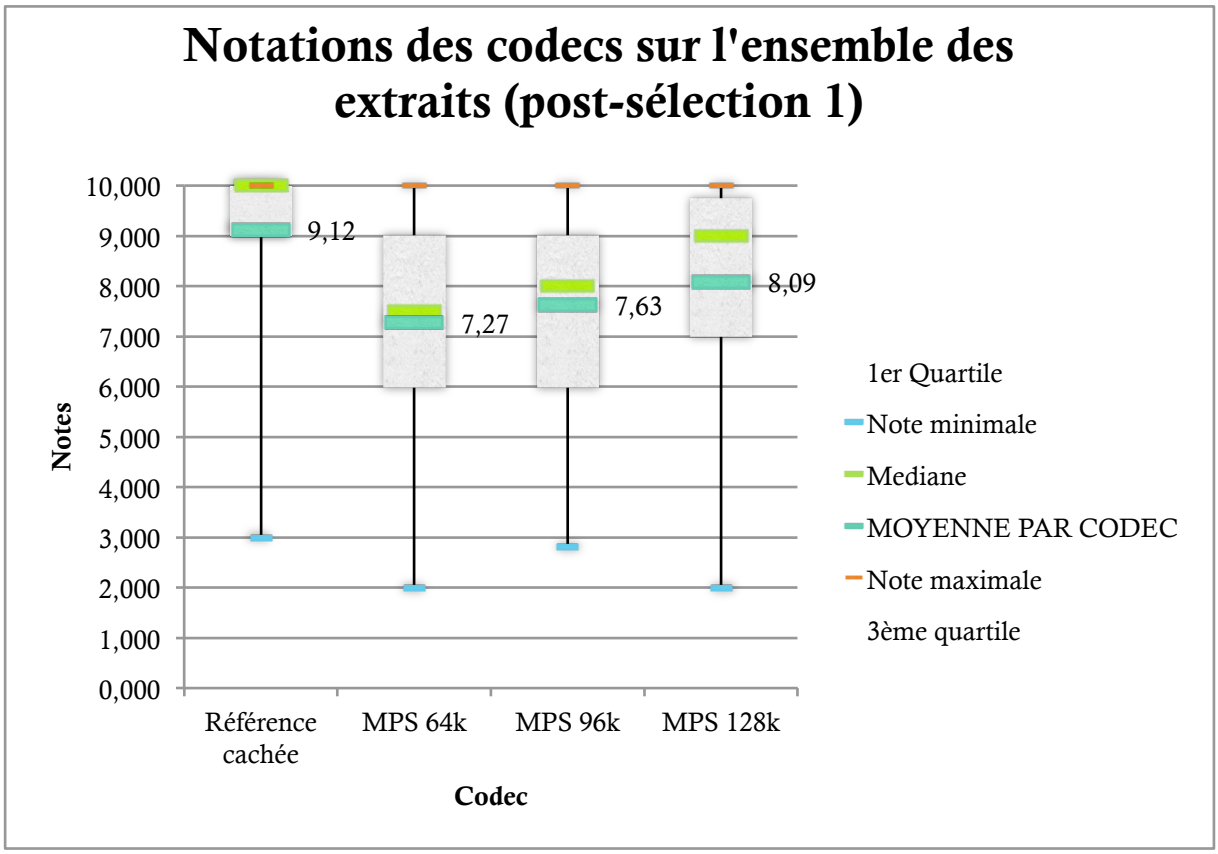


Figure 88 : Notations des codecs sur l'ensemble des extraits après la 1ère post-sélection

Malgré la première post-sélection, on s'aperçoit que la référence cachée a quand même obtenue au moins une note égale à 3/10. Une deuxième étape de post-sélection consiste à éliminer les candidats retenus, ayant jugé une ou plusieurs références cachées à moins de 5/10 : j'élimine donc trois candidats supplémentaires.

Il me reste donc quatorze candidats.

Tableau 9 : Notations des extraits après la 2ème post-sélection

Notations des codecs (tous extraits confondus) après la 2ème post-sélection				
CODEC	Référence cachée	MPS 64k	MPS 96k	MPS 128k
MOYENNE PAR CODEC	9,36	7,52	7,80	8,28
ECART-TYPE (σ_{N-1})	1,102	1,920	1,812	1,718
Note minimale	5,000	2,000	4,000	2,000
1er Quartile	9,000	6,000	7,000	7,000
Médiane	10,000	8,000	8,000	9,000
3ème quartile	10,000	9,000	9,000	9,875
Note maximale	10,000	10,000	10,000	10,000

La 2ème post-sélection permet d'obtenir une meilleure moyenne pour la référence cachée : 9,36/10, et un classement par ordre croissant des trois codecs en fonction du débit, mais ici encore les moyennes des codecs sont très proches : 7,52/10 pour le codec à 64 kbits/seconde, 7,80/10 pour le codec à 96 kbits/seconde et 8,28/10 pour le codec à 128 kbits/seconde, avec des écarts-types plus resserrés qu'avant la post-sélection. Néanmoins, on observe toujours que tous les codecs ont eu au moins une note égale à 10/10, ce qui signifie que le participant a évalué ce codec d'une qualité identique à celle de l'original.

Une comparaison multiple par paires a de nouveau été réalisée : les résultats sont les mêmes que précédemment. Les échantillons de la variable « référence cachée » se démarquent, mais le codec MPEG Surround à 96 kbits/seconde ne se distingue pas des deux autres codecs.

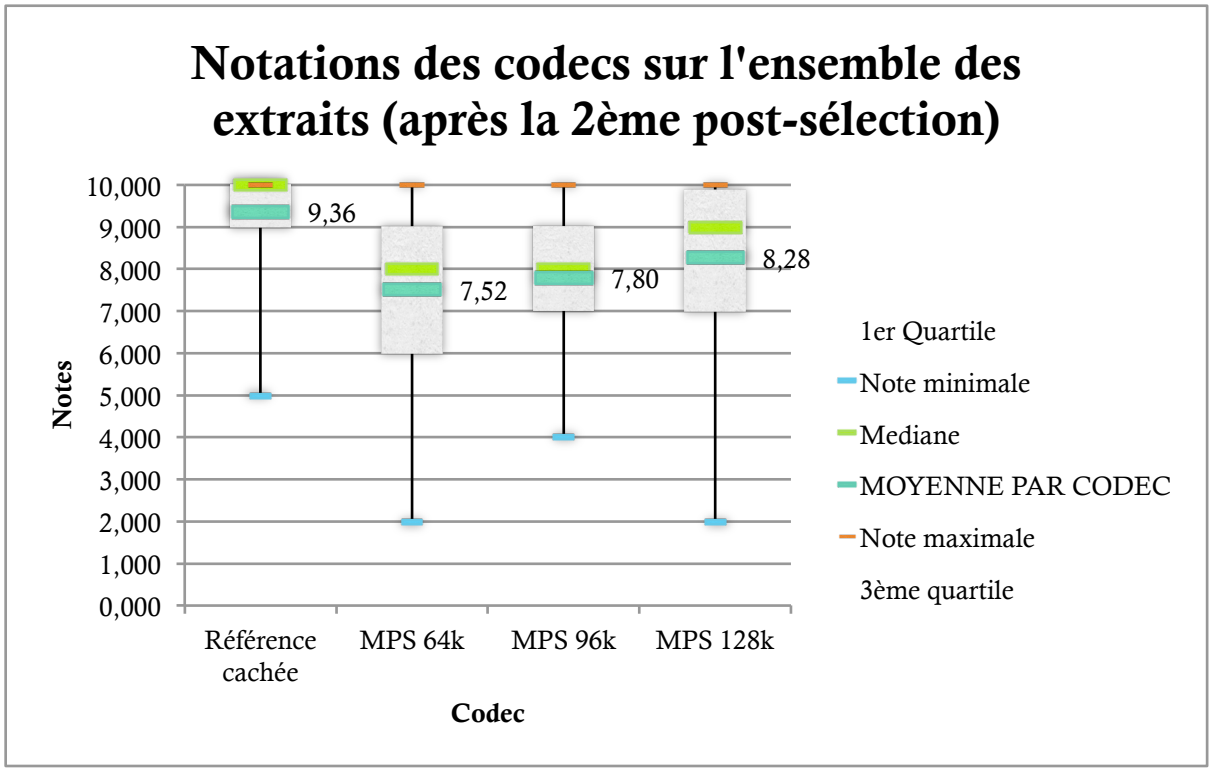


Figure 89 : Moyennes des codecs tous extraits confondus (après la 2ème post-sélection)

Je réalise un nouveau test de Shapiro-Wilk pour déterminer si les variables suivent une loi normale après post-sélection. Le test de Shapiro-Wilk rejette l'hypothèse de normalité pour toutes les variables, néanmoins le graphique montre que la distribution des trois codecs MPEG Surround est tout de même proche de la droite. Je vais donc tenter tout de même un calcul d'intervalle de confiance de 95% en appliquant la loi de Student à N-1 = 14 -1 = 13 degrés de liberté, ne connaissant pas de manière de le calculer pour une distribution non normale. Je suis consciente que les résultats ne seront pas optimaux, mais donneront une première idée.

Selon la table de Student, pour un intervalle de confiance de 95% à 13 degrés de liberté, on applique un coefficient de 2,016.

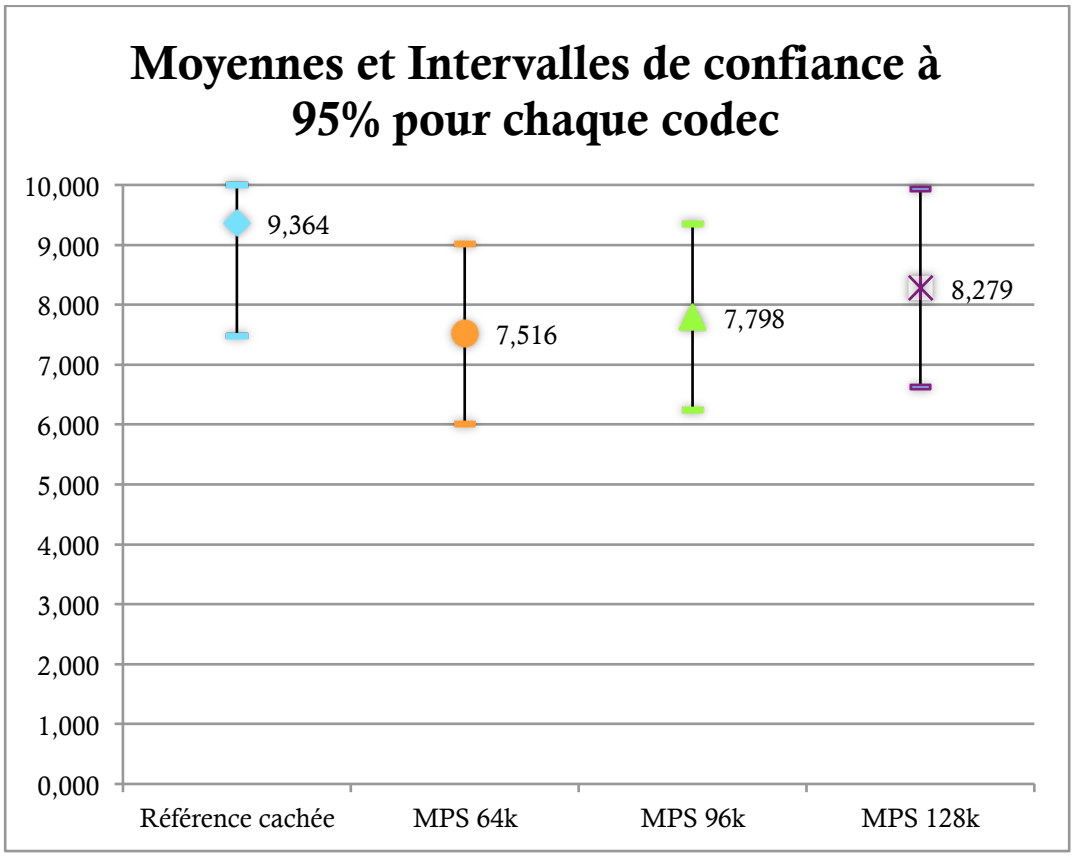


Figure 90 : Moyennes et Intervalles de confiance à 95% pour chaque codec

On remarque que les intervalles de confiance se superposent et que les moyennes des trois codecs sont très proches : les moyennes de chaque codec ne sont donc pas significativement différentes. Afin de s'assurer quel codec est le plus performant, il faudrait mener une autre campagne de tests, avec des extraits de programmes télévisés plus discriminants, et réaliser une pré-sélection des candidats pour s'assurer qu'ils ont l'habitude des tests d'écoute.

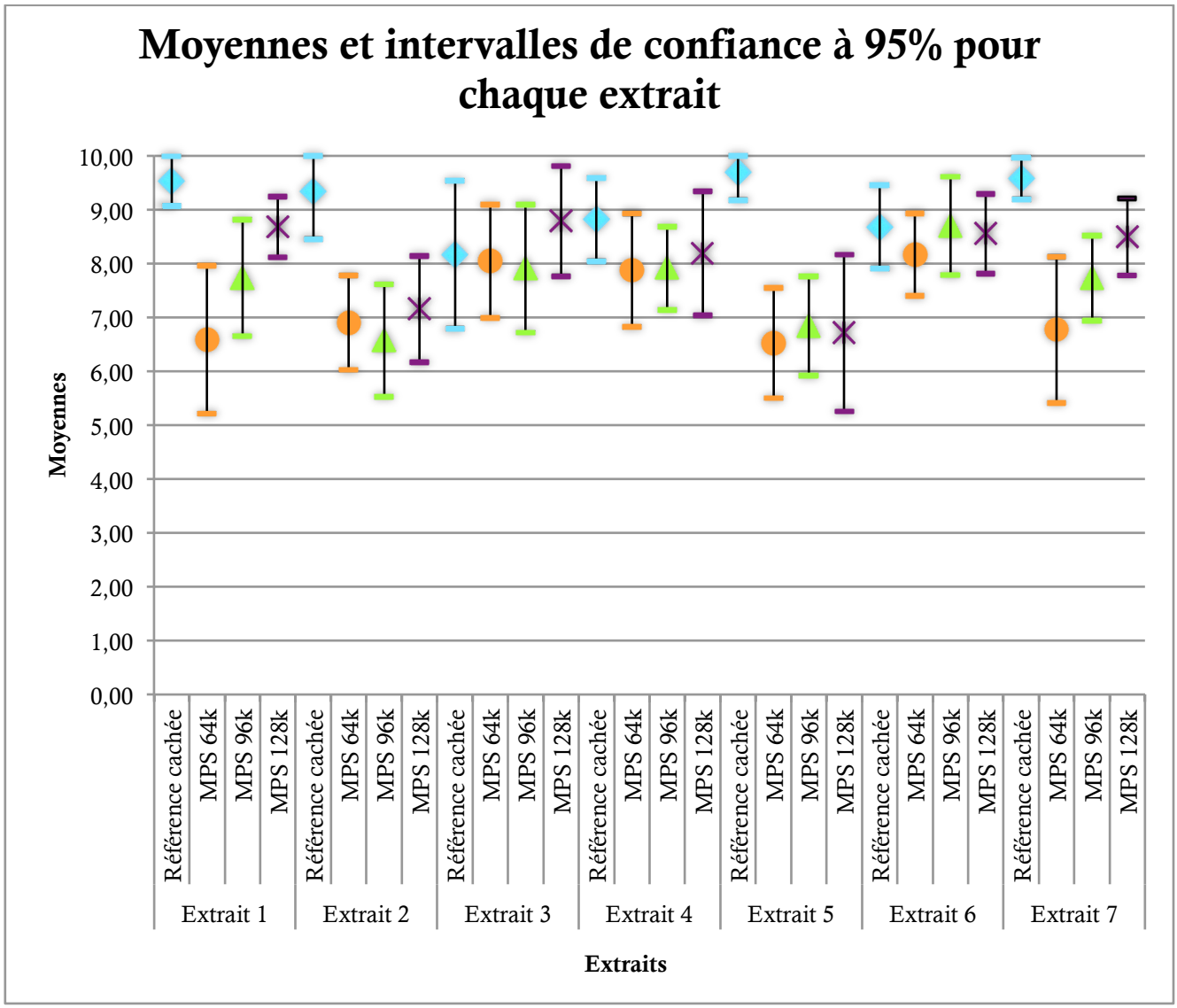


Figure 91 : Moyennes et Intervalles de confiance pour chaque codec et chaque extrait

Pour chaque extrait, les intervalles de confiance se superposent : les moyennes ne sont donc pas significativement différentes. En revanche, on remarque que les moyennes de l'extrait 1 sont tout de même espacées, et le codec HE-AAC + MPEG Surround à 128 kbits/seconde est bien identifié comme le meilleur.

La référence cachée de l'extrait 3 a quasiment la même moyenne que les codecs à 64 kbits/seconde et 96 kbits/seconde : cet extrait illustre la performance du codec sur des contenus audio constitués essentiellement de voix, dans le canal central.

L'extrait 2 a obtenu une moyenne de 9,34/10 pour la référence cachée, moyenne significativement différente des trois encodages.

L'extrait 5 est l'extrait qui obtient la meilleure moyenne pour la référence cachée, tandis que les trois encodages sont notés plus sévèrement que pour d'autres extraits : cet extrait a été l'un des plus discriminants, avec les applaudissements de l'extrait 2.

L'extrait 6 obtient une moyenne quasiment égale pour la référence cachée et les deux encodages au débit les plus élevés : seul le débit à 64 kbits/seconde obtient une moyenne légèrement inférieure : le codage MPEG Surround est donc performant sur cet extrait, puisque les candidats ne différencient pas la référence cachée des autres codecs.

L'extrait 7 démontre une notation croissante des signaux en fonction du débit : les candidats ont donc dans l'ensemble réussi à percevoir des détails dans les fonds d'air.

Cette post-sélection permet d'évaluer de façon plus fiable les différents encodages. Néanmoins, la comparaison des moyennes de chaque codec tous extraits confondus révèle une faible différence entre le codage à 64 kbits/seconde et celui à 96 kbits/seconde, qui sont évalués dans la partie supérieure du critère de qualité « Bon », tandis que le codage à 128 kbits/seconde est considéré comme « excellent ».

4.2.4. Principaux Artefacts Détectés

L'encodage HE-AAC + MPEG Surround induit plusieurs types d'artefacts : tout d'abord, le timbre est dégradé, les très hautes fréquences (supérieures à 12 kHz) sont coupées et induisent alors moins de brillance. Ce défaut est particulièrement sensible sur les passages musicaux avec percussions ou certains effets sonores. Néanmoins, ces fréquences là n'étant plus perçues par toutes les oreilles, ce défaut n'a pas toujours été soulevé. Sinon, on remarque parfois une spatialisation plus incertaine, qui serait particulièrement sensible avec des sources sonores sèches et ponctuelles, mais qui se dilue dans une acoustique réverbérante (par exemple, dans les extraits d'opéra, ce phénomène ne se remarque quasiment pas car on a majoritairement des tutti d'orchestre avec de la réverbération. Ce défaut serait davantage perceptible sur un instrument solo dans une acoustique moins réverbérante). Ce défaut est toutefois perçu sur les applaudissements,

qui n'ont pas la même position selon le débit d'encodage utilisé. Concernant la spatialisation, on remarque aussi parfois une réduction de l'image stéréophonique, ceci est particulièrement flagrant sur le début de l'extrait 5, où la musique originale avait une belle largeur stéréophonique, et cette même musique se retrouve beaucoup plus étriquée et centrée sur les signaux encodés et décodés à 64 et 96 kbits/seconde. Il s'agit là d'un artefact, qu'il faut néanmoins nuancer : si ce défaut est perceptible dans ces conditions là, avec les enceintes gauche et droite écartées de … mètres, je doute que ce soit aussi flagrant sur un système home-cinéma placé dans un salon, où les enceintes seront cette fois écartées de un à deux mètres maximum, où plusieurs personnes écouteront sans forcément être au sweet-spot. De plus, certaines personnes peuvent préférer une image stéréophonique plus resserrée et ne pas être gênées par ce défaut.

Le canal LFE semble plutôt bien reproduit, quelque soit le débit utilisé.

Les candidats ont tous été impressionnés par la performance du codage MPEG Surround, et ce quelque soit le débit. Même s'ils ont entendu des faibles différences, certains ont eu des difficultés à évaluer ces différences, et manquaient d'une référence inférieure : un signal d'ancre les aurait peut-être aidés. De plus, la consigne les a un peu déroutés : leur demander de donner une appréciation globale du signal était une question trop vague pour certains, et auraient préféré évaluer le timbre, la dynamique, la localisation, etc. Mais le but de mon test était avant tout d'obtenir une impression globale, et si je leur avais demandé d'évaluer trois ou quatre paramètres pour chaque extrait, ils auraient eu besoin de plus de temps par extrait, et auraient probablement fini par noter complètement au hasard tous les paramètres. Avec une seule question, je suis consciente que certains se sont davantage concentrés sur tel ou tel paramètre, mais entre tous, j'obtiens une appréciation globale de chaque signal.

Globalement, sur ces sept extraits, le codage MPEG Surround associé au codeur HE-AAC est plutôt performant, puisque rares sont les personnes à donner des notes inférieures à 5, au débit le plus faible (64 kbits/seconde). Certes, ce codage produit des

artefacts, mais ils restent acceptables, et perceptibles seulement par des oreilles averties, dans des conditions d'écoute particulières.

4.3. CONCLUSION

Il est difficile d'émettre une conclusion objective et fiable avec les résultats obtenus. Les résultats ne sont pas suffisamment différenciés pour conclure quant au meilleur rapport qualité/débit. Des tendances se dessinent, globalement la qualité du codage augmente avec le débit, mais les différences ne sont pas marquées. De plus, la difficulté à retrouver la référence cachée est un phénomène à ne pas négliger. Évidemment, dans l'absolu le codage au plus haut débit, à 128 kbits/seconde semble le plus performant, mais si la différence de qualité avec des plus faibles débits n'est pas significative, je ne peux pas vraiment conclure.

Il faut cependant noter que parmi les extraits choisis, il y avait globalement peu de contenus dans les canaux arrière, et les signaux ne sont pas forcément les plus difficiles (un extrait de musique jazz, nécessitant beaucoup de précision et ayant moins de réverbération, un extrait d'applaudissements seuls, etc. auraient peut-être éprouvé davantage le codec). Mais il faut retenir que ce codage, avec ses qualités et ses défauts, semble adapté à la majorité des programmes télévisés (talk-shows, fictions).

Néanmoins, je peux déjà tirer des conclusions quant au protocole choisi : le choix des extraits était-il judicieux ? Les extraits étaient-ils assez significatifs ? Incontestablement, on se doute que des extraits plus discriminants, comme des applaudissements seuls, des percussions dans une acoustique sèche, des castagnettes, des signaux avec des variations de dynamique instantanée très importantes, des signaux ultra-compressés, des signaux comportant davantage de contenus dans les canaux arrière, ou encore des signaux totalement artificiels et n'ayant aucune ressemblance avec un contenu télévisuel , tous ces signaux auraient peut-être permis une différenciation plus évidente, comme dans le test décrit chapitre 4, paragraphe 3, page 90. Néanmoins, le choix

d'évaluer des programmes télévisés encodés était délibéré puisque d'autres tests avaient déjà été effectués, et qu'il était important de tester de vrais contenus télévisés.

On peut aussi remettre en question le protocole : aurait-il été préférable d'effectuer seulement des tests en double aveugle à triple stimuli comme décrit dans la recommandation ITU-R BS.1116 car protocole plus adapté à de faibles dégradations ?

En effet, la méthodologie MUSHRA est préconisée pour des dégradations moyennes, et utilisée dans tous les tests du codage MPEG Surround publiés à l'AES, mais vues les premières réactions des candidats, les codecs semblent créer des dégradations plus faibles que celles auxquelles on s'attendait. De plus, son principal avantage, outre le gain de temps, est son offre de comparaison directe, qui permet ainsi de comparer chaque signal à l'original, mais aussi d'évaluer chaque signal par rapport à un autre de façon relative. Par exemple, le candidat compare les signaux A, B, C et D au signal de référence. Il détermine que le signal B est la référence cachée car identique, et perçoit des différences pour les autres. MUSHRA lui permet alors de comparer les signaux A, C et D entre eux, afin de définir lequel est le meilleur, ou tout du moins lequel est le plus acceptable et le plus proche du signal de référence, lequel est celui de plus mauvaise qualité, etc. Avec un test en double aveugle à triple stimuli, le candidat compare à chaque fois un codec et un signal de référence cachée à un signal de référence : premièrement, il a plus de chances de trouver la référence cachée au hasard. De plus, admettons qu'au premier essai, le codec inconnu est le codec à 64 kbits/seconde et au deuxième essai le codec à 96 kbits/seconde. Lors du premier essai, il peut détecter la référence cachée et noter l'autre signal à 7/10 par exemple, lui trouvant certains défauts. A l'essai suivant, il peut réussir encore à détecter la référence cachée, et évaluer l'autre signal à 5/10, sans pouvoir évaluer si ce codec là est meilleur ou moins bon que celui de l'essai précédent, car sa mémoire auditive ne lui permet pas la comparaison. Un test en double aveugle à triple stimuli ne permettrait donc pas de comparer trois codecs (ou plus) entre eux, mais seulement de les comparer à l'original.

En revanche, la méthodologie MUSHRA conseille d'utiliser parmi les signaux à évaluer, outre un signal de référence cachée, un signal d'ancre, qui est souvent un signal original filtré à 3,5 ou 7 kHz. Dans mon test, on avait décidé de s'affranchir de ce signal

d'ancre, afin de ne pas ajouter un signal de plus à évaluer. Ce fut peut-être une erreur : cette ancre n'aurait probablement pas évité les erreurs de détection de la référence cachée, mais aurait peut-être servi de base inférieure de notation aux participants, et aurait aussi permis d'évaluer la capacité des candidats à différencier les dégradations des différents encodages. L'absence de phase d'entraînement a peut-être aussi été un handicap pour certains participants puisqu'ils ne connaissaient pas les types de défauts produits par ce codage, mais l'ordre aléatoire des séquences a minimisé cet inconvénient. Enfin, on peut se demander si on ne s'est pas fourvoyé concernant la notion d'auditeur « expert » : est-ce que des étudiants ont une écoute suffisamment aiguisée ? Est-ce que des assistants son, des techniciens son et des personnes dans le domaine de l'ingénierie pratiquent suffisamment d'écoutes critiques pour être considérées comme « expertes » ? Aurait-il fallu convier à ce test seulement des candidats ayant l'habitude de ce genre de tests pour obtenir des résultats plus fiables ? Ce dernier critère aurait limité le nombre de candidats, mais les résultats auraient peut-être été plus homogènes ?

Enfin, on peut se demander si la présence de l'image n'a pas influencé les candidats et minimisé les différences perçues ? Évidemment, l'image a probablement un impact mais pour tester le rendu de la spatialisation, il était important de le comparer avec l'image. De plus, beaucoup de candidats ont naturellement fermé les yeux pour se concentrer davantage.

Toutes ces réflexions montrent à quel point organiser un test perceptif est loin d'être aisé, il faut tenir compte de bon nombre de paramètres, avec des contraintes de temps, de planning, de budget, de disponibilité des candidats, etc.

Une deuxième semaine de tests perceptifs était programmée afin de compléter les premiers résultats : les tests menés ensuite sont plus informels mais complètent l'approche de ce codage.

5. TESTS COMPLÉMENTAIRES:

Pour cette deuxième série de tests, j'envisageais plusieurs hypothèses : mener une nouvelle série de tests sur un système home-cinéma cette fois pour évaluer le rendu du codage dans un milieu proche de l'écoute grand public, réaliser les mêmes tests avec d'autres extraits de programmes télévisés (avec ou sans image d'ailleurs, et surtout des extraits ayant plus d'informations dans les canaux arrières), tester la compatibilité stéréophonique en comparant le downmix stéréo du MPEG Surround avec un downmix plus traditionnel du signal 5.1 original, ou encore tester la compatibilité binaurale, etc. Par faute de temps, il a fallu renoncer à certains tests et ceux que l'on a privilégiés ont été menés avec moins de participants, de façon informelle, mais dans le but de compléter les données déjà obtenues.

Les résultats de la première série de tests menés dans un laboratoire avec un écran de 4 mètres (et donc des enceintes gauche et droite éloignées de 4 mètres), et un système audio presque neutre, ont montré des différences relativement faibles entre les différents codages. Il s'avère donc inutile de mener ces mêmes tests sur un système home-cinéma de moyenne gamme, voire même de haute gamme, qui lisseraient sans doute encore davantage les différences.

En ce qui concerne la compatibilité binaurale, le plugin Fraunhofer Pro-Codec ne propose un encodage binaural avec une seule HRTF, qu'on ne peut modifier, et trois simulations de pièces : une pièce sèche (Dry Room), une pièce de moyenne taille (Living Room) et une grande salle type cinéma (Cinema). De plus, cela nécessite un autre encodage. Cette option n'est donc pas intéressante, puisqu'on sait que la reproduction binaurale basée sur des HRTF standards fonctionne moyennement. En revanche, le laboratoire dispose de l'appareil Realiser A8 de Smyth, qui permet de prendre les empreintes de nos HRTF en cinq minutes. Une fois l'empreinte enregistrée, l'appareil encode en temps réel le son binaural du signal 5.1 reçu. J'aurais pu mener quelques tests pour avoir le ressenti de quelques personnes de l'écoute de l'audio 5.1 encodé en MPEG Surround, puis décodé et ré-encodé en binaural, mais il aurait surtout été intéressant de

mener ce genre de tests sur une tablette, pour confronter l'impression de téléspectateurs regardant une vidéo sur une tablette en écoutant du son binaural : ce test aurait nécessité trop de temps de préparation, ou il aurait fallu faire déplacer les participants deux fois : une fois pour réaliser l'empreinte de leurs HRTF, puis une fois pour le test. Dans le temps qu'il me restait, c'était difficile à mettre en œuvre.

En revanche, il nous a paru plus pertinent de tester la compatibilité stéréophonique du MPEG Surround et de refaire quelques tests d'écoute sur le même système 5.1 avec des extraits ayant davantage de contenus sonores dans les canaux arrières. Le test sera donc divisé en deux parties : une première partie de tests avec de nouveaux extraits encodés en MPEG Surround associé au codeur HE-AAC aux trois mêmes débits utilisés dans le premier test, puis décodés en 5.1, et une deuxième partie de test d'écoute en stéréophonie au casque pour comparer la qualité du downmix MPEG Surround à un downmix plus traditionnel.

5.1 Protocole de la Deuxième Série de Tests

Pour ce deuxième test, je disposais des mêmes programmes télévisés, dans lesquels j'ai sélectionné de nouveaux extraits. Afin de compléter les premières données obtenues, ce deuxième test s'est déroulé en son seul, sans image. J'ai sélectionné trois nouveaux extraits pour la première partie d'écoute en 5.1, d'une durée de 40 secondes environ : un extrait de l'opéra *L'Italiana in Algeri*, un extrait du téléfilm *Jusqu'à l'enfer* et un extrait du documentaire *La vie moderne* de Raymond Depardon, extraits qui comportent plus de contenus sonores dans les canaux arrière et l'extrait de téléfilm avec des effets sonores. J'ai aussi sélectionné quatre autres extraits pour la partie d'écoute au casque, dont un extrait utilisé lors de la première série de tests, et qui avait été particulièrement dégradé par le codage MPEG Surround.

Tableau 10 : Description des extraits du 2ème test

Extrait	Programme	Type	Durée	Description
Ext 11	*L'Italiana in Algeri* 2ème acte	Opéra filmé	39s20i	Fin d'un air (Acte 2 - scène 11) : soprano + chœur et orchestre, puis applaudissements.
Ext 12	*Jusqu'à l'enfer* De Denis Malleval	Téléfilm	43s06i	Scène de cauchemar avec des effets sonores (sound design), une réminiscence d'une voix.
Ext 13	*La vie moderne* De Raymond Depardon	Documentaire	38s05i	Scène dans un salon, le paysan regarde la messe à la télévision, et répond aux questions du réalisateur. Briquet, et le son de la télévision très frontal, se déplace progressivement vers les canaux arrière (champ : téléviseur ; contre-champ : paysan qui regarde la télé)
Ext 21	*L'Italiana in Algeri* 2ème acte	Opéra filmé	33s03i	Fin d'un air « Cavatina » (Acte 2 – scène 3) : ténor + orchestre, puis applaudissements et début de la scène suivante au piano.
Ext 22	*Jusqu'à l'enfer* De Denis Malleval	Téléfilm	42s05i	Scène avec des effets sonores, puis déplacement des aboiements du chien de la gauche entre avant et arrière et le centre.
Ext 23	*Jusqu'à l'enfer* De Denis Malleval	Téléfilm	53s24i	Scène avec alternance de la maquette du train et du train réel. Musique (image stéréo large), sons de la maquette et du véritable train, sound design, pas sur les cailloux, klaxons du train. Canal LFE chargé. (Même scène qu'au test 1)
Ext 24	*La vie moderne* De Raymond Depardon	Documentaire	37s14i	Interview en extérieur d'une jeune agricultrice, en journée. Ambiance calme, peu d'arrière.

La première partie du test est identique à la première série de tests : l'écoute comparative se fait sur un système 5.1, avec un niveau d'écoute réglé à 70 dB. Pour chaque extrait, le candidat dispose du signal original 5.1 appelé référence, et de quatre stimuli nommés « A », « B », « C » et « D », parmi lesquels se cache une copie du signal de

référence (nommée « référence cachée »), et trois stimuli encodés en MPEG Surround associé au codeur HE-AAC à trois débits différents : 64 kbits/seconde, 96 kbits/seconde et 128 kbits/seconde. La notation est la même : une échelle continue de 0 à 10, graduée par pas de 1. La référence cachée doit obtenir la note de 10. La durée est limitée à cinq minutes par extrait.

Pour la deuxième partie du test, le but était de comparer quatre downmix stéréophoniques différents au casque : les downmix MPEG Surround associés au codeur HE-AAC aux trois débits utilisés précédemment (64 kbits/seconde, 96 kbits/seconde et 128 kbits/seconde), avec un downmix « traditionnel ». Ces quatre stimuli sont répartis de façon aléatoire sur les pistes nommées « A », « B », « C » et « D ». Afin d'évaluer la qualité audio du downmix créé lors de l'encodage en MPEG Surround, il a fallu l'extraire du flux MPEG-4 (que ProTools ne sait pas lire), et pour cela on l'a converti à l'aide du logiciel FFMpeg en fichier audio stéréophonique en PCM (.wav), à 48 kHz et en 24 bits (cf Annexe D pour la ligne de commande, le logiciel FFMpeg, libre de droits, ne dispose pas de décodeur MPEG Surround). Le downmix « traditionnel » est en fait un downmix stéréo créé manuellement à partir du fichier original en 5.1, avec ces coefficients là :

Tableau 11 : Coefficients de pondération des canaux pour le downmix manuel

	L	R	C	LFE	Ls	Rs
L	1.0	0.0	0.707	0.0	0.707	0.0
R	0.0	1.0	0.707	0.0	0.0	0.707

En fait, il s'avère que le plugin Sonnox Fraunhofer Pro-Codec appliquent ces mêmes gains pour réaliser le downmix stéréo en MPEG Surround, mais agit ensuite de façon à préserver l'énergie sonore, afin d'éviter l'annulation ou la sommation de certaines fréquences, une atténuation de 3 dB est aussi appliquée, et un limiteur est encore ajouté afin d'éviter des saturations ponctuelles. Cette atténuation de 3 dB est réversible lors du décodage MPEG Surround.

Ici, on ne dispose pas de signal de référence, on a donc décidé de comparer ces quatre downmix les uns par rapport aux autres, et de leur donner une note comprise entre

0 et 10, sur la même échelle que précédemment. Néanmoins, ici, aucun stimulus n'est obligé d'avoir la note de 10.

Figure 92 : Realiser A8 de Smyth

Le test est alors réalisé sur le même principe, le candidat peut choisir d'écouter le downmix qu'il souhaite et faire les comparaisons de son choix par l'intermédiaire des boutons « solo » de la surface de contrôle Command 8. L'appareil Realiser A8 de Smyth est alors utilisé comme amplificateur de casque. Afin d'écouter un signal stéréophonique natif, il suffit d'utiliser le mode « stereo mix-down » avec les paramètres suivants :

Canal gauche : 1.0 x L ; 0.0 x R ; 0.0 x C, 0.0 x LFE, 0.0 x Ls, 0.0 x Rs

Canal droit : 0.0 x L ; 1.0 x R ; 0.0 x C, 0.0 x LFE, 0.0 x Ls, 0.0 x Rs

Afin d'éviter une saturation en entrée de l'amplificateur de casque, j'utilise une piste stéréophonique auxiliaire, avec un niveau de -11 dB, dans laquelle j'envoie chaque piste stéréophonique du test, qui contient les stimuli nommés « A », « B », « C » et « D », cette piste Auxiliaire sort dans la sortie stéréophonique principale, et l'écoute principale est « mutée ». Le casque utilisé est un casque ouvert DT 990PRO de Beyerdynamic, confortable pour l'écoute.

Une fois les extraits audio sélectionnés, j'ai réalisé une session ProTools, avec cinq pistes audio 5.1, quatre pistes audio stéréophoniques et une piste stéréophonique auxiliaire, et j'ai réparti de façon aléatoire les différents stimuli. Par manque de temps et peu de participants, on a jugé peu utile de créer une session par participant

Il est évident que l'ordre du test, c'est-à-dire commencer par un test sur un système d'écoute en 5.1, puis enchaîner avec un test au casque, a un impact sur le ressenti et la notation des extraits. J'ai choisi de réaliser le test dans ce sens afin de permettre aux candidats novices (ceux qui n'ont jamais écouté de l'audio encodé en MPEG Surround) d'apprécier les qualités et les défauts du codage, afin de les comparer à ceux des downmix

stéréophoniques lors de l'écoute au casque. De plus, puisque ce test est réalisé avec moins de participants, il n'a pas de véritable valeur statistique, le but étant surtout de dessiner certaines tendances, et de répondre à la demande de certains participants, qui lors du premier test, étaient curieux d'écouter du MPEG Surround au casque.

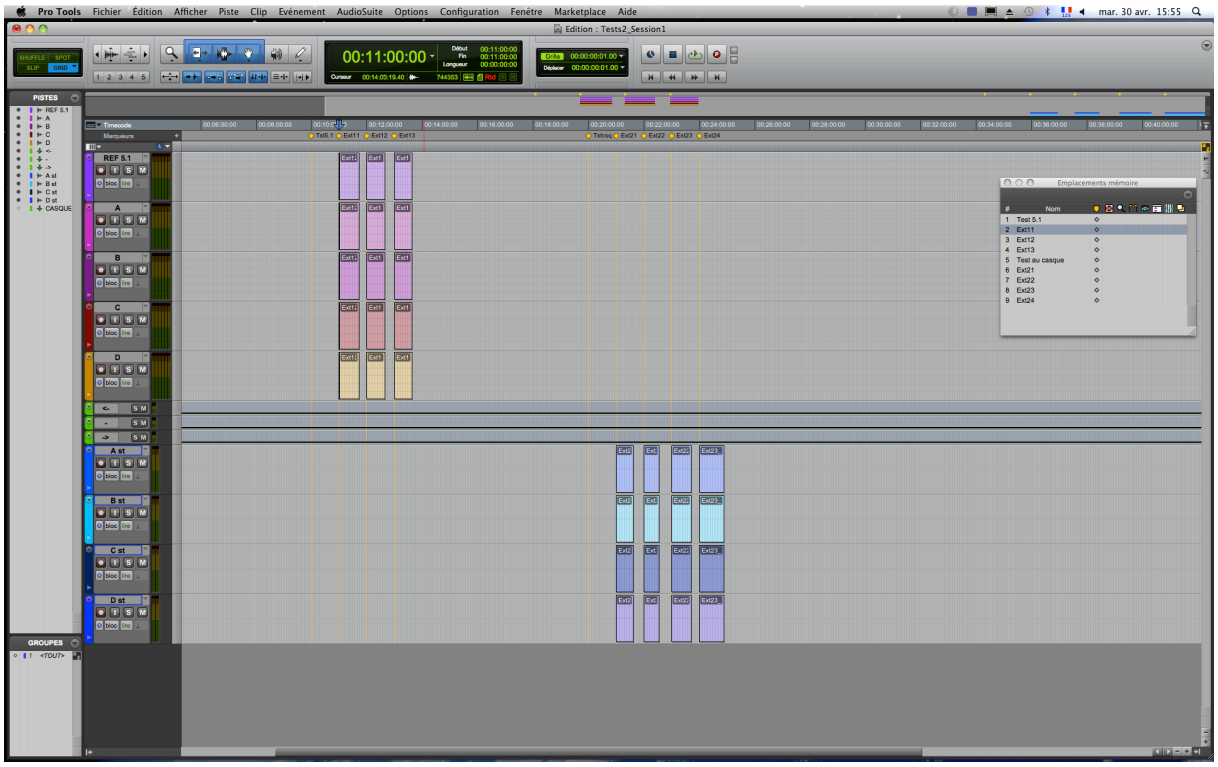
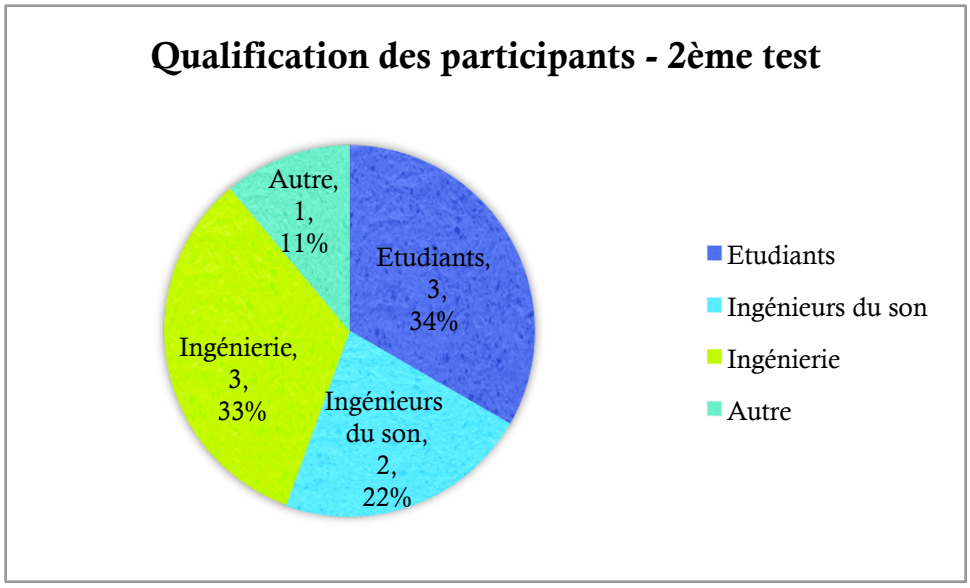


Figure 93 : Capture d'écran : Session Test 2

5.2 PARTICIPANTS

Ce deuxième test a été mené avec neuf participants, dont trois étudiants en formation son, deux ingénieurs du son, trois personnes travaillant dans les domaines de l'ingénierie ou responsables audio, et une personne non professionnelle du son mais musicienne. Parmi eux, six candidats avaient participé à la première série de tests, et avaient, par conséquent, déjà écoutés du MPEG Surround.

Figure 94 : Qualification des participants du 2ème test



5.3 RÉSULTATS

Les résultats statistiques obtenus sont donnés à titre purement indicatif, puisque le nombre de participants est trop faible. Toutefois, il s'agit d'en tirer certaines tendances.

Parmi les neuf candidats, j'ai choisi d'en éliminer un, qui faute d'entendre des différences dans la deuxième partie et sans point de repère, a préféré de ne pas noter certains extraits. Il reste donc huit candidats.

Pour la première partie du test en 5.1, la référence cachée obtient une très bonne moyenne : 9,23/10. En revanche, l'extrait 12, la référence cachée n'a été retrouvée que par quatre candidats sur les huit participants. En revanche, elle a été facilement retrouvée pour les deux autres extraits. Le codec HE-AAC + MPEG Surround à 128 kbits/seconde se détache des deux autres et obtient la meilleure moyenne : 8,5/10, la différence avec les autres codecs est d'ailleurs plus marquée que lors du premier test. Le codec à 64 kbits/seconde obtient tout de même une moyenne de 7,21/10, qui correspond à une bonne qualité. En revanche, l'écart-type est très étendu.

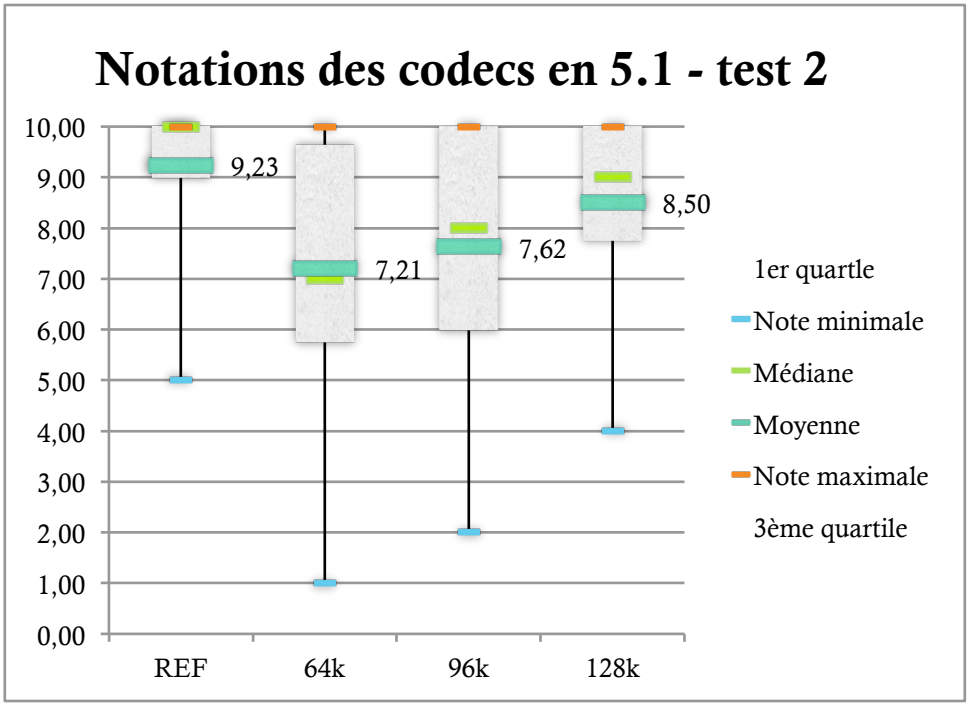


Figure 95: Notations des extraits 5.1 du 2ème test

Tableau 12 : Notations des extraits 5.1 lors du 2ème test

Test 2 – Écoute 5.1 – Tous extraits confondus				
CODEC	Référence Cachée	MPS 5.1 64k	MPS 5.1 96k	MPS 5.1 128k
MOYENNE	9,233	7,213	7,625	8,500
ECART-TYPE (σ_{N-1})	1,439	2,490	2,337	1,581
Note minimale	5,00	1,00	2,00	4,00
1er quartile	9,00	5,75	6,00	7,75
Médiane	10,00	7,00	8,00	9,00
3ème quartile	10,00	9,63	10,00	10,00
Note maximale	10,00	10,00	10,00	10,00

Pour la deuxième partie du test au casque, on observe que le downmix réalisé de façon traditionnelle obtient la meilleure moyenne, seulement pour deux extraits sur quatre. D'ailleurs, le downmix MPEG Surround à 64 kbits/seconde obtient la meilleure moyenne des downmix MPEG Surround, tous extraits confondus. De plus, les écarts-types sont étendus. Cela démontre alors que les différences perceptibles sont faibles, et donc non significatives, sur ces extraits là en tout cas.

Les participants ont même mentionné que les défauts étaient moins avérés au casque que sur le système 5.1. La compatibilité stéréophonique du MPEG Surround est donc assurée. Le downmix stéréophonique donne une image stéréophonique plus large, avec davantage de profondeur, mais les trois encodages MPEG Surround s'en tirent bien, avec une moyenne de 7,39/10, pour le codec HE-AAC + MPEG Surround à 96 kbits/seconde, qui récolte la moyenne la plus faible.

Tableau 13 : Notations des extraits 2.0 lors du 2ème test

Test 2 – Écoute 2.0 : Notations des codecs				
CODEC	Stéréo Downmix	MPS 2.0 64k	MPS 2.0 96k	MPS 2.0 128k
MOYENNE	8,294	8,019	7,391	7,869
ECART-TYPE (σ_{N-1})	1,541	1,770	2,210	1,667
Note minimale	4,00	3,00	2,00	3,00
1er quartile	7,00	7,00	6,00	7,00
Médiane	8,00	8,00	7,50	8,00
3ème quartile	10,00	10,00	9,00	9,00
Note Maximale	10,00	10,00	10,00	10,00

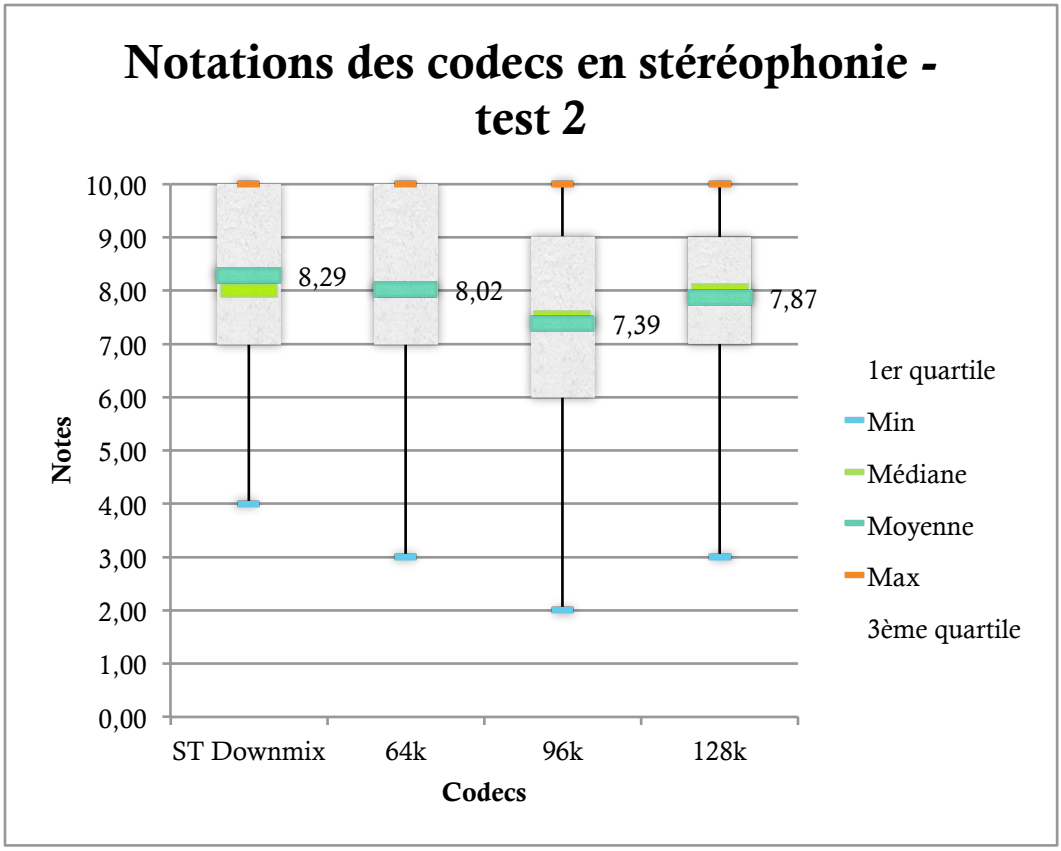


Figure 96 : Notations des extraits stéréophoniques - test 2

5.4. Remarques Concernant le Protocole

Certains participants m'ont fait remarquer que la deuxième partie du test, au casque, était difficile car ils n'avaient pas de référence, la meilleure note obtenue par le codec qu'ils préféraient n'était donc pas forcément la note maximale de l'échelle. De plus, certains pensaient pouvoir déceler des défauts du codage MPEG Surround beaucoup plus facilement au casque, et il s'avère que les défauts ne sont pas plus flagrants au casque que sur enceintes. C'est un bon point de plus pour les codecs.

Les résultats sont donc à considérer avec prudence. Il était difficile de choisir quel signal stéréophonique serait signal de référence : le downmixing stéréophonique est matricé, tandis que les autres subissent un matriçage ainsi qu'un filtrage, et la version 2.0 du programme est un véritable mix stéréophonique. Néanmoins, je suis consciente

qu'outre le fait que j'avais un nombre insuffisant de participants, l'absence de signal repère a faussé les résultats. Cependant, ce deuxième test a été mené par curiosité, pour essayer de voir si les artefacts entendus sur la version reconstruite en 5.1 sont autant présents sur le downmix, et d'après ces premiers résultats, les downmix MPEG Surround sont tout à fait satisfaisants.

6. CONCLUSION

En confrontant les analyses des deux séries de tests perceptifs, je peux donc conclure que le codage MPEG Surround, associé au codage principal HE-AAC, est très performant, même au débit le plus faible. Les différences de notation entre les débits ne sont pas assez marquées, il m'est donc difficile de conclure quant au meilleur choix de débit : un débit de 128 kbits/seconde est théoriquement le plus performant, néanmoins le débit de 96 kbits/seconde obtient une moyenne très proche. Ce codage a évidemment des défauts et produit des artefacts, tels qu'une dégradation du timbre dans les hautes fréquences essentiellement, une réduction de l'image stéréophonique, une localisation parfois moins précise. Néanmoins, ces artefacts produits sur des extraits de programmes télévisés n'ont pas été perçus par tous les candidats, même des professionnels du son.

En outre, on aurait pu penser que les extraits d'opéra, avec un spectre très riche, auraient été les plus impactés par le codage. Et en fait, les passages musicaux n'ont pas forcément été les plus dégradés : bien qu'ils aient subi une dégradation dans le haut du spectre et une légère réduction de l'image stéréophonique, seules les oreilles jeunes et/ou habituées à l'écoute de musique classique ont perçu ces défauts. Ce sont les applaudissements qui ont éprouvé le plus le codec, tous débits confondus.

On peut donc en conclure que ce codage est tout à fait acceptable pour la majorité des programmes télévisés, à savoir des talk-shows, des magazines, et de la fiction. Cependant, il conviendrait de s'en assurer en réalisant d'autres tests, plus élaborés, avec d'autres extraits, et de véritables « experts », qui ont l'habitude des tests d'écoute et des comparaisons de codecs, afin de confirmer ces premiers résultats obtenus.

Il pourrait aussi être intéressant de réaliser des tests en écoute binaural en visionnant un programme sur une tablette afin d'observer la réaction des candidats.

CHAPITRE 6 : MPEG SURROUND ET RECOMMANDATION R128

Depuis le 1er Janvier 2012, tous les nouveaux programmes diffusés sur des chaînes de télévision française doivent respecter la recommandation R128. Désormais, depuis le 1er Janvier 2013, tous les programmes, qu'ils aient été produits avant ou après cette date, doivent suivre cette recommandation.

On se demande donc comment le format MPEG Surround agit sur les paramètres tels que l'intensité sonore du programme, l'intensité sonore à court-terme, la valeur de crête vraie ou encore la distribution statistique de l'énergie sonore.

1. NAISSANCE DE LA RECOMMANDATION R128

La recommandation EBU R128 est née en 2011, elle instaure un nouveau système de mesure du niveau sonore des programmes diffusés à la télévision, qui raisonne en terme de « loudness », c'est-à-dire en terme de niveau moyen plutôt qu'en niveaux crêtes. L'organisme américain ATSC[39] définit le terme « loudness » de la façon suivante : « A perceptual quantity ; the magnitude of the physiological effect produced when a sound stimulates the ear ». En français, on peut définir ce terme comme « intensité sonore

[39] ATSC ou Advanced Television Systems Committee, Inc. est un organisme américain équivalent à l'EBU (European Broadcasting Union) en Europe ou encore la CST (Commission Supérieure Technique de l'image et du son) en France

ressentie », puisque les algorithmes de mesures tiennent compte des caractéristiques de l'audition humaine afin de donner une mesure du niveau subjectif d'intensité sonore.

Cette recommandation R128 a été créée dans le but d'homogénéiser les niveaux sonores entre différents programmes télévisés d'une même chaîne, ainsi qu'entre différentes chaînes. En effet, il n'y a rien de plus désagréable que de subir de fortes variations de niveaux sonores lors du passage d'un programme à une publicité, ou lors d'une scène d'action dans un film alors que les personnages chuchotaient dans la séquence précédente. Cette dynamique trop étendue obligeait le téléspectateur à agir en permanence sur le niveau sonore, afin de pouvoir comprendre les passages dialogués tout en gardant un niveau maximal acceptable.

En France, tous les diffuseurs ont accepté cette recommandation, et désormais ils refusent des prêts-à-diffuser si les paramètres « loudness » ne respectent pas cette recommandation.

Les mixeurs de longs métrages détestent réaliser des versions télés, d'autant plus maintenant avec cette recommandation, ils ont l'impression de détruire leur travail et de renoncer aux intentions qu'ils voulaient donner. Or, si les conditions de diffusion et d'écoute au cinéma sont plutôt idéales puisque l'audio multicanal n'est pas compressé et l'écoute est régulièrement contrôlée, il n'est pas de même pour la télévision, dont le débit de l'audio est réduit pour la diffusion, puis le son est souvent écouté sur les haut-parleurs de l'écran de télévision, dans un environnement plus ou moins bruyant, selon le lieu et l'isolation de la pièce.

De plus, les chaînes se livraient à une bataille sans fin, à savoir qui diffuserait le plus fort, en compressant au maximum tous les sons afin d'augmenter le niveau moyen sans dépasser le niveau maximal toléré, alors égal à - 9 dB Full Scale.

Les conditions d'écoute étant très diverses, entre les personnes qui regardent la télévision en définition standard ou en haute définition, il était essentiel d'établir une réglementation afin de proposer des programmes qui ont retrouvé une dynamique acceptable tout en ménageant l'intelligibilité des dialogues et une homogénéité de niveaux entre différents programmes.

2. Les Paramètres de la Recommandation R128

La recommandation ITU-R BS.1770 définit les méthodes de mesures des niveaux loudness, pour des signaux monophoniques, stéréophoniques et multicanaux 5.0, le canal LFE n'étant pas considéré pour les mesures, tandis que la recommandation R128 détermine les valeurs cibles que ces paramètres doivent atteindre. La FICAM[40], le HDForum[41] et les éditeurs de la TNT ont signé la recommandation technique RT 017 de la CST[42], qui précise les caractéristiques des signaux audio et vidéo des Prêts-À-Diffuser.

Figure 97 : Logo de la recommandation R128

2.1. La Pondération K

Afin d'effectuer des mesures de loudness, un nouvel algorithme de mesures a été créé, qui nécessite une pondération appelée pondération K : cette pondération est réalisée par un double-filtre de second ordre, de pente douce. Le pré-filtre reproduit la fonction de transfert d'une tête humaine, qui diffracte les signaux sonores. Le second filtre, nommé R2LB (Revised Low Frequency), simule les caractéristiques physiologiques moyennes de l'oreille humaine. Lorsque l'on cascade ces deux filtres, on obtient un filtre dit de pondération K, qui tient compte de la perception auditive humaine. Cette pondération est issue de nombreux tests psycho-acoustiques menés par différents chercheurs.

---

[40] FICAM : Fédération des Industries du Cinéma, de l'Audiovisuel et du Multimédia, est une organisation syndicale patronale qui regroupe plus de 170 entreprises spécialisées dans les métiers de l'image et du son.

[41] HDForum : LE HDForum regroupe une cinquantaine de sociétés, afin d'optimiser les chaînes audiovisuelles de production, de diffusion et de restitution sur le parc de téléviseur haute définition.

[42] CST : Comission Supérieure Technique de l'Image et du Son, est une association de professionnels de l'audiovisuel, chargés de veiller à la qualité de la chaîne de production et de diffusion des images et du son, pour le cinéma, la télévision ou tout autre média.

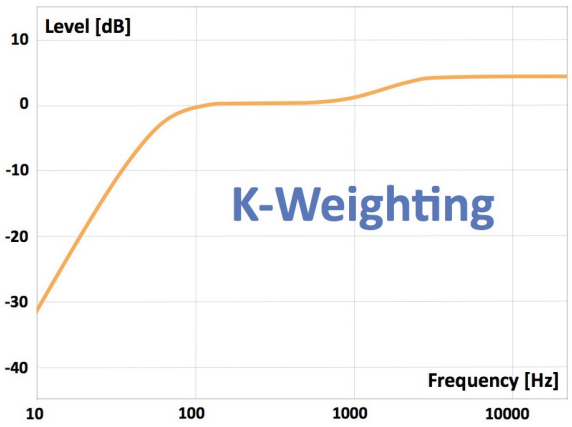


Figure 98 : Courbe de pondération K

2.2. La Sommation

Pour mesurer l'intensité sonore ressentie d'un programme, on réalise une pondération K sur les cinq canaux principaux (le canal LFE n'est pas pris en compte dans les calculs), on prend la valeur efficace moyenne de ces cinq canaux séparément, auxquelles on applique un coefficient de pondération, puis on somme toutes ces valeurs et on obtient l'intensité sonore du programme.

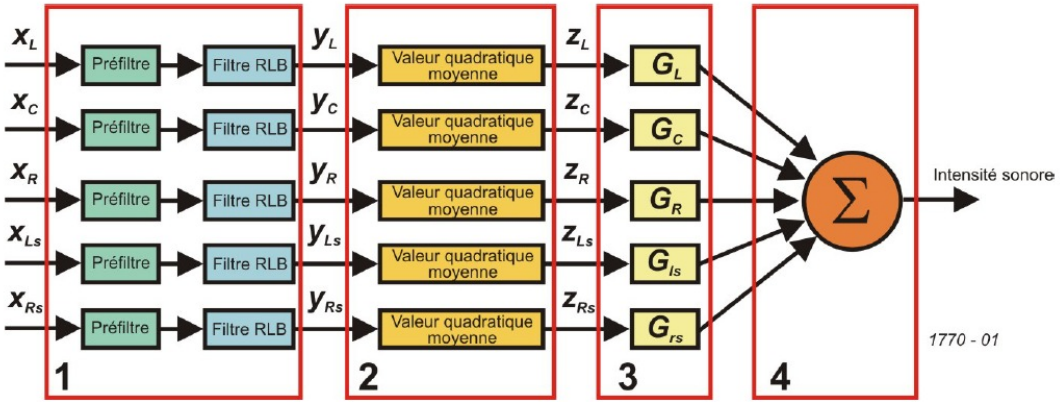


Figure 99 : Schéma de principe de l'algorithme de mesure de l'intensité sonore ressentie d'un programme

En pratique, les trois canaux frontaux (gauche, centre, droit) ne sont pas pondérés. En revanche, on applique un gain de 1,5 dB aux deux canaux arrière. Ceci vient du fait

que les canaux arrière semblent perçus plus forts qu'ils ne le sont réellement. Les valeurs efficaces moyennes pondérées de chaque canal sont alors sommées selon cette équation :

$$Intensité\ sonore = -0{,}691 + 10 \log_{10} \sum_{i}^{N} G_i \cdot z_i$$

Avec G_i le coefficient de pondération de chaque canal, et z_i la valeur efficace moyenne de l'énergie, mesurée dans un intervalle de temps T.

Le temps d'intégration T n'est pas précisé, on utilise donc la même formule de calcul quelque soit la durée du programme (une publicité de quelques secondes ou un long métrage de deux heures). On obtient alors le niveau d'intensité sonore du programme, aussi appelé Programme Loudness, qui s'exprime, soit en valeur absolue en Loudness Unit relatif à une échelle dite Full Scale (LUFS), soit en valeur relative en Loudness Unit (LU). Cette unité exprime donc l'intensité sonore d'un programme pondéré K. L'échelle LUFS est graduée par pas de 1 LU, tel que 1 LU = 1 dB, et 0 LU = - 23 LUFS. Désormais, chaque programme télévisé, quelque soit sa duré, doit atteindre la valeur cible d'intensité sonore de -23 LUFS +/-1 LU.

2.3. Le Niveau de Crête Vrai ou True Peak

Auparavant, les crêtes-mètres numériques (ou QPPM Quasi Peak Programme Meter) fonctionnaient en échantillons, c'est-à-dire qu'ils ne fournissaient que les valeurs des échantillons numériques, tandis que les crêtes-mètres analogiques avaient un temps de réponse d'environ 10 ms, et étaient donc incapables de détecter des crêtes dont la durée était inférieure à ce temps de réponse. De plus, le niveau de crête vrai d'un signal est parfois supérieur à la valeur maximale de l'échantillon car elle peut intervenir entre deux échantillons. Et lors de la conversion numérique-analogique ou lors de traitements numériques, l'extrapolation des échantillons peut alors recréer ces crêtes et provoquer des saturations.

Afin de mesurer un véritable niveau de crête, qu'on appellera crête vraie, un nouvel algorithme de mesure a été mis au point, incluant notamment un quadruple sur-échantillonnage, afin d'obtenir une forme d'onde beaucoup plus précise et de s'assurer qu'aucune autre crête n'est présente entre deux échantillons.

Puisque cette mesure de crête est plus fiable, on pourrait donc imposer une limitation à -1 dB True Peak. Mais dans sa recommandation technique, la CST prend une marge et préconise une valeur de crête vraie maximale de -3 dBTP.

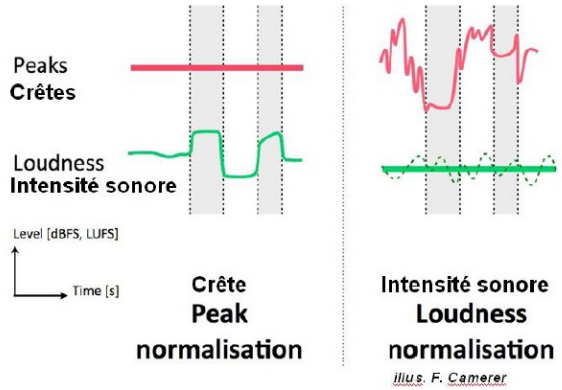


Figure 100 : Comparaison d'une normalisation par crête ou par loudness

Cette nouvelle recommandation permet alors de regagner de la dynamique, tout en homogénéisant les niveaux de diffusion, comme le montre la figure 100.

2.4. Les Trois Indicateurs d'Intensité Sonore

Trois modes de mesure de l'intensité sonore existent.

Le mode « Momentary Loudness » ou mode momentané, noté « M », mesure l'intensité sonore du programme sur une fenêtre glissante de 400 ms, qui pourrait s'apparenter à celle d'un vu-mètre. Elle est peu utilisée en France.

Le mode « Short-term Loudness » ou mode court-terme, noté « S » mesure l'intensité sonore du programme sur une fenêtre carrée glissante de trois secondes, avec un chevauchement entre deux fenêtres de 66% au minimum, soit une mesure toutes les secondes environ. Ce mode est surtout utilisé pour contrôler le niveau sonore des dialogues, la valeur cible à atteindre étant -23 LUFS à +/- 7 LU, c'est-à-dire que les cris

ne doivent pas dépasser un niveau de -16 LUFS tandis que les murmures ne doivent pas moduler en dessous de -30 LUFS. Cette préconisation garantit l'intelligibilité de la parole, mais elle est souvent difficile à respecter, et nécessite une surveillance permanente de la part du mixeur.

Enfin le mode « Gated Integrated Loudness » ou mode « moyenne », noté « I », est une mesure du niveau d'intensité sonore avec seuil, réalisée sur l'ensemble du programme, et doit atteindre la valeur cible de -23 LUFS à +/- 1 LU. La valeur loudness de chaque canal est mesurée sur une fenêtre glissante de 400 ms, avec 70% de chevauchement, soit une mesure toutes les 100 millisecondes. Cette mesure tient compte d'un seuil absolu (tous les blocs dont le loudness est inférieur à -70 LUFS sont ignorés) et d'un seuil relatif (une fois les blocs précédents supprimés, un calcul de loudness avec les blocs restants est réalisé, on place un seuil relatif 10 LU en dessous de cette valeur, on supprime tous les blocs dont l'intensité sonore est inférieure à ce seuil, et on calcule à nouveau l'intensité sonore du programme, en s'affranchissant des blocs supprimés). Les seuils admettent qu'on élimine les blocs dont l'énergie sonore est trop faible, de façon absolue et relativement au niveau moyen du loudness mesuré.

2.5. La Distribution Statistique de l'Énergie Sonore ou Loudness Range (LRA)

Le dernier paramètre primordial de cette recommandation est la distribution statistique de l'énergie sonore ou loudness range (LRA). Cette grandeur est une mesure statistique de la variation du loudness d'un programme entier, qui permet d'évaluer la dynamique du signal, en s'affranchissant des blocs ponctuels dont l'intensité sonore est très forte ou trop faible, et qui perturberaient la mesure. L'algorithme effectue une mesure de l'intensité sonore en mode « court-terme », les blocs dont la valeur loudness est inférieure à -70 LUFS sont ignorés, les autres sont stockés dans une table, tandis qu'on calcule une nouvelle fois l'intensité sonore d'un programme qui serait uniquement composé que des blocs mémorisés. On place un seuil relatif 20 LU en dessous de cette

valeur, et on évalue la distribution des niveaux de modulation. La valeur de loudness range est définie comme la différence des fenêtres de 95% et de 10% de la distribution.

La recommandation R128 et la recommandation RT 017 préconisent une valeur de LRA comprise entre 5 et 20 LU. Si la distribution statistique est trop étendue, c'est-à-dire si la valeur de LRA est supérieure à 20 LU, les diffuseurs refusent le prêt-à-diffuser, car cela signifie que le programme est trop dynamique.

2.6. VALEURS CIBLES

Tous les programmes, quelque soit leur durée (d'une publicité de quelques secondes à une version télé de long métrage de deux heures) doivent donc respecter les valeurs cibles imposées par la Commission Supérieure Technique de l'Image et du Son, à savoir :

- ♪ L'intensité sonore ressentie du programme doit être égale à -23 LUFS à +/-1 LU,
- ♪ L'intensité sonore à court-terme doit être égale à -23 LUFS à +/- 7 LU,
- ♪ Le niveau de crête vraie doit être inférieur à – 3 dB True Peak,
- ♪ La distribution statistique de l'énergie sonore doit être comprise entre 5 et 20 LU.

Pour les mesurer, il existe une vaste gamme d'outils : des plugins, tels que le Dolby Media Meter ou le SLM 128, des appareils externes tels que le Touch Monitor 9 de RTW, ou encore des libraires à implémenter en ligne de commande. Chaque outil est plus ou moins adapté à chaque situation, on préférera un appareil externe pour une émission en direct, tandis que les plugins sont plus adaptés en postproduction.

3. L'Impact du Codage MPEG Surround sur les Mesures de Loudness

En 2013, puisque tous les programmes télévisés sont soumis à une recommandation en matière de loudness, il paraissait inévitable de réaliser une petite étude concernant l'impact du codage MPEG Surround sur les paramètres tels que l'intensité sonore du programme, le niveau de crête vraie et la distribution statistique de l'énergie sonore. Dans l'idéal, le codage MPEG Surround ne devrait pas influencer le niveau d'intensité sonore, ni la distribution statistique de l'énergie sonore, en revanche une légère différence sur la valeur de crête vraie paraît inéluctable.

On a donc encodé et décodé plusieurs programmes télévisés complets, pour mesurer avant encodage, après encodage et après décodage ces trois principaux paramètres.

3.1. Protocole de Mesures

A l'aide d'un programme d'analyse et de mesures loudness, on a mesuré l'intensité sonore du programme, le niveau de crête vraie et la distribution statistique de l'énergie sonore sur les signaux originaux 5.1 et stéréophonique intégraux des trois programmes utilisés lors des tests, puisque la recommandation R128 s'applique à un programme entier. J'ai alors encodé et décodé ces trois programmes entiers en MPEG Surround, associé au codage HE-AAC, aux trois débits utilisés lors des tests, c'est-à-dire en 64 kbits/seconde, en 96 kbits/seconde et en 128 kbits/seconde. Afin de garantir des mesures équitables, les mesures ont été menées sur les programmes intégraux, auxquels étaient ajoutés deux bips d'une durée d'une image et quatre secondes de silence numérique (non prises en compte dans les mesures).

Les programmes entiers encodés en MPEG Surround dans un flux MPEG-4 et contenant le downmix stéréophonique ont été convertis en un fichier stéréophonique en

format PCM linéaire (.wav), en 48 kHz en 24 bits avec le logiciel libre de droits FFMPEG[43], qui ne sait pas décoder le MPEG Surround, puisque soumis à une licence.

Une fois la conversion réalisée, je disposais donc des trois programmes intégraux utilisés dans les tests, à savoir le téléfilm *Jusqu'à l'Enfer* de Denis Malleval, le documentaire *La vie moderne* de Raymond Depardon et l'opéra *L'Italiana in Algeri* filmé à l'Opéra National de Lorraine, en 5.1 et en stéréophonie original (en 48kHz – 24 bits), en version stéréophonique après encodage MPEG Surround associé au codeur HE-AAC aux trois débits (64 kbits/seconde, 96 kbits/seconde et 128 kbits/seconde), et en version 5.1 reconstituée après décodage. A l'aide de la librairie « libebur128 » libre de droits, qui implémente la mesure des paramètres loudness selon la recommandation R128, et d'un programme de commandes, on a alors mesurer les trois paramètres suivants : intensité sonore du programme, niveau de crête vraie et distribution statistique de l'énergie sonore.

La comparaison des mesures entre le programme 5.1 original et le programme reconstruit en 5.1 après encodage permettra de déterminer si le codec agit sur les paramètres loudness préconisés par le recommandation R128 (et si le débit utilisé agit aussi). Les mesures sur le downmix stéréo MPEG Surround permettent de vérifier la compatibilité stéréophonique/5.1 au niveau des paramètres de cette même recommandation.

3.2. TAUX DE COMPRESSION

Afin de mesurer les paramètres loudness des fichiers entiers des trois programmes, j'ai donc dû les encoder et les décoder. J'ai alors pu relever la taille des différents fichiers et calculer les taux de compression.

[43] La ligne de commande utilisée pour la conversion du flux MPEG-4 en un fichier stéréophonique est disponible en annexe D, pages 212 et 213 .

Tableau 14 : Taux de compression du codec MPEG Surround + HE-AAC à 64 kbits/seconde

	PCM (original)	MPS+HE-AAC à 64 kbits/seconde			Taux de compression	
	5.1 .wav	2.0 .mp4	2.0 .wav	5.1 .wav	2.0 .wav /.mp4	5.1/ downmix
Jusqu'à l'enfer (1h35 57s)	5,157 Go	47,1 Mo	1,658 Go	5,157 Go	35,20	109,49
La vie moderne (1h23 08s)	4,467 Go	40,8 Mo	1,437 Go	4,467 Go	35,22	109,49
L'Italiana in Algeri (2h31 44s)	8,154 Go	74,5 Mo	2,622 Go	8,154 Go	35,19	109,45

Tableau 15 : Taux de compression du codec MPEG Surround + HE-AAC à 96 kbits/seconde

	PCM (original)	MPS + HE-AAC à 96 kbits/seconde			Taux de compression	
	5.1 .wav	2.0 .mp4	2.0 .wav	5.1 .wav	2.0 .wav /.mp4	5.1/ downmix
Jusqu'à l'enfer (1h35 57s)	5,157 Go	70,2 Mo	1,658 Go	5,157 Go	23,62	73,46
La vie moderne (1h23 08s)	4,467 Go	60,8 Mo	1,437 Go	4,467 Go	23,63	73,47
L'Italiana in Algeri (2h31 44s)	8,154 Go	111 Mo	2,622 Go	8,154 Go	23,62	73,46

Tableau 16 : Taux de compression du codec MPEG Surround + HE-AAC à 128 kbits/seconde

	PCM (original)	MPS+HE-AAC à 128 kbits/seconde			Taux de compression	
	5.1 .wav	2.0 .mp4	2.0 .wav	5.1 .wav	2.0 .wav /.mp4	5.1/ downmix
Jusqu'à l'enfer (1h35 57s)	5,157 Go	93,2 Mo	1,658 Go	5,157 Go	17,79	55,33
La vie moderne (1h23 08s)	4,467 Go	80,7 Mo	1,437 Go	4,467 Go	17,81	55,35
L'Italiana in Algeri (2h31 44s)	8,154 Go	147,4 Mo	2,622 Go	8,154 Go	17,79	55,32

On peut remarquer que le taux de compression des fichiers encodés en MPEG Surround ne dépend que du codeur principal et du débit utilisés. La durée du fichier ou son contenu sont négligeables.

La colonne 2.0 .wav/.mp4 représente le taux de compression d'un flux MPEG-4 contenant le downmix stéréophonique MPEG Surround et les données de spatialisation par rapport à un fichier stéréophonique non compressé.

La colonne 5.1/downmix représente le taux de compression d'un flux MPEG-4 contenant le downmix stéréophonique MPEG Surround et les données de spatialisation par rapport au fichier 5.1 original en PCM linéaire.

Pour un débit de 64 kbits/seconde, le codec MPEG Surround associé au codeur HE-AAC permet d'atteindre un taux de compression quasiment égal à 110 : 1. C'est une réduction de débit considérable, d'autant plus en comparaison de la qualité très acceptable de ce codec, bien qu'il y ait des artefacts.

Quand on augmente le débit, le taux de compression diminue un peu, on a un ratio de 73,5 : 1 à 96 kbits/seconde, et 55,3 : 1 à 128 kbits/seconde, ce qui permet déjà une très bonne économie de bande-passante, tout en garantissant une excellente qualité.

Quel codec avec des taux de compression semblables d'une qualité excellente pourrait venir concurrencer le MPEG Surround? A ma connaissance, aucun.

3.3. RÉSULTATS DE MESURES LOUDNESS

Extrait	Téléfilm *Jusqu'à l'enfer*, de Denis Malleval								
Encodage MPEG Surround	PCM (.wav ou .bwf)		HE-AAC + MPS à 64 kbits/s		HE-AAC + MPS à 96 kbits/s		HE-AAC + MPS à 128 kbits/s		
Audio	5.1 original	LtRt original	Enc. MPS (2.0)	Déc. MPS (5.1)	Enc. MPS (2.0)	Déc. MPS (5.1)	Enc. MPS (2.0)	Déc. MPS (5.1)	Valeurs cibles (rappel)
Intensité sonore du programme (LUFS)	-24,9	-24,1	-28	-25,1	-28	-25	-27,9	-25	-23 LUFS à +/-1 LU
Niveau de crête vraie (dB TP)	1,8	-1,8	-1	0,3	1,1	1,9	-1,8	1	<-3 dB TP
Distribution statistique de l'énergie sonore (LU)	18,4	18,5	18,3	18,3	18,4	18,4	18,4	18,4	5 LU < LRA < 20 LU

Tableau 17 : Mesures loudness du téléfilm et du documentaire

Extrait	Documentaire *La vie moderne*, de Raymond Depardon								
Encodage MPEG Surround	PCM (.wav ou .bwf)		HE-AAC + MPS à 64 kbits/s		HE-AAC + MPS à 96 kbits/s		HE-AAC + MPS à 128 kbits/s		
Audio	5.1 original	2.0 original	Enc. MPS (2.0)	Déc. MPS (5.1)	Enc. MPS (2.0)	Déc. MPS (5.1)	Enc. MPS (2.0)	Déc. MPS (5.1)	Valeurs cibles (rappel)
Intensité sonore du programme (LUFS)	-27,7	/	-29,8	-27,8	-29,8	-27,7	-29,7	-27,7	-23 LUFS à +/-1 LU
Niveau de crête vraie (dB TP)	-1,7	/	-3,7	-1,6	-3,8	-1,6	-3,7	-1,6	<-3 dB TP
Distribution statistique de l'énergie sonore (LU)	17,4	/	18,3	17,4	18,3	17,5	18,4	17,5	5 LU < LRA < 20 LU

Extrait	Opéra *L'Italiana in Algeri,* de Gioacchino Rossini Intégrale								
Encodage MPEG Surround	PCM (.wav ou .bwf)		HE-AAC + MPS à 64 kbits/s		HE-AAC + MPS à 96 kbits/s		HE-AAC + MPS à 128 kbits/s		
Audio	5.1 original	2.0 original	Enc. MPS (2.0)	Déc. MPS (5.1)	Enc. MPS (2.0)	Déc. MPS (5.1)	Enc. MPS (2.0)	Déc. MPS (5.1)	Valeurs cibles (rappel)
Intensité sonore du programme (LUFS)	-23,6	-23,6	-25,4	-23,7	-25,4	-23,5	-25,4	-23,6	-23 LUFS à +/-1 LU
Niveau de crête vraie (dB TP)	-1,8	-3,1	-3,5	-2,3	-3,2	-2	-3,4	-2,1	<-3 dB TP
Distribution statistique de l'énergie sonore (LU)	15	14,6	15,1	15	15,1	15	15	15	5 LU < LRA < 20 LU

Tableau 18 : Mesures loudness de l'opéra "L'Italiana in Algeri"

L'opéra *L'Italiana in Algeri* étant en deux actes, on a mesuré les paramètres loudness sur chaque acte, et sur les deux actes réunis.

3.4. ANALYSE DES RÉSULTATS

Tout d'abord, on peur noter que l'intensité sonore du programme original ainsi que le niveau de crête vraie ne correspondent pas aux valeurs cibles fixées par la recommandation R128 pour le téléfilm et le documentaire. Ces deux programmes ont été mixés avant 2012, et cette recommandation n'était alors pas en vigueur. L'opéra a quant à lui été mixé courant 2012, et respecte donc les valeurs cibles (à l'exception du niveau de crête vraie pour le programme en 5.1).

On peut remarquer que la valeur de l'intensité sonore des programmes décodés en 5.1 est quasiment identique à celle du programme original en 5.1 (à +/- 0,2 LU), et la valeur de la distribution statistique de l'énergie sonore est aussi conservée (à +/- 0,1 LU), ce qui est tout à fait satisfaisant. En revanche, le niveau de crête vraie varie, jusqu'à 1,5

LU, ce qui est considérable, mais on pouvait s'y attendre. Le seul programme qui conserve son niveau de crête vraie est le documentaire, car la dynamique instantanée est limitée.

En revanche, on remarque qu'entre le signal 5.1 original et les downmix stéréophoniques issus de l'encodage en MPEG Surround, les valeurs de loudness ne correspondent pas, à l'exception de la distribution statistique de l'énergie sonore, qui se maintient à +/- 1 LU. Les downmix stéréophoniques ont une intensité sonore de 1,7 à 3,1 LU inférieure à celle du signal 5.1 original. De même, le niveau de crête vraie des downmix MPEG Surround est de 1 à 3 LU inférieur à celui du signal 5.1 original. Cette constatation est problématique à l'heure où tous les programmes télévisés diffusés en France doivent respecter les valeurs imposées par la recommandation CST/FICAM RT-017, recommandation signée par tous les diffuseurs français et qui reprend les valeurs préconisées par la recommandation R128, alors que nos mesures montrent que la compatibilité 5.1/stéréophonie n'est pas assurée en matière de loudness. Ceci représente alors un frein majeur à l'implémentation de ce codage pour la diffusion de programmes télévisés.

En fait, après quelques recherches, ce problème vient de l'encodage en MPEG Surround réalisé par le plugin Fraunhofer Pro-Codec de Sonnox. La 1ère étape de cet encodage MPEG Surround consiste à réaliser un downmix stéréophonique, tout en extrayant les données de spatialisation. Les gains utilisés pour réaliser ce downmix sont les suivants (ce sont les mêmes que le downmix traditionnel réalisé manuellement pour la deuxième série de tests perceptifs) :

	L	R	C	LFE	Ls	Rs
L	1.0	0.0	0.707	0.0	0.707	0.0
R	0.0	1.0	0.707	0.0	0.0	0.707

Tableau 19 : Coefficients pour le downmix stéréophonique

A ces gains, ils ajoutent un algorithme pour préserver l'énergie sonore, en évitant la sommation ou l'annulation de certaines fréquences. Afin d'éviter des saturations dues aux sommations, un facteur constant de -3 dB est appliqué, ce facteur est alors réversible

lors du décodage. Enfin, un limiteur est ajouté pour pallier aux saturations des signaux très forts. Ceci explique alors pourquoi le signal décodé retrouve des paramètres loudness quasiment égaux à ceux du signal 5.1 d'origine, et pourquoi on observe une intensité sonore et un niveau de crête vraie du downmix environ 3 LU en-dessous des valeurs du signal original en 5.1.

Pour remédier à ce désagrément, on pourrait envisager d'insérer la valeur d'intensité sonore dans les métadonnées du downmix stéréophonique (en HE-AAC ou en AAC) ou alors ajouter une étape à la fin de l'encodage afin que le signal stéréophonique obtienne les mêmes paramètres loudness que le signal 5.1 original. Néanmoins, ces options ne sont pas encore disponibles dans l'implémentation de ce plugin.

4. CONCLUSION

Étant donné l'incompatibilité stéréophonique/5.1 concernant la recommandation R128, il paraît difficile que le codage MPEG Surround soit utilisé dans la diffusion de flux télévisés dans un avenir proche, d'autant plus que Fraunhofer est pour l'instant le seul fabricant à proposer un encodeur/décodeur MPEG Surround. Néanmoins, si leurs algorithmes d'encodage venaient à évoluer et à intégrer une modification, afin que le downmix stéréophonique ait la même intensité sonore que le signal décodé en 5.1, il pourrait alors être intéressant de réévaluer ce codage et sa possible utilisation.

CONCLUSION

Les différentes expérimentations menées au cours de ces quatre mois ont été très intéressantes : elles m'ont permis de découvrir les caractéristiques du codage MPEG Surround, de comprendre comment l'algorithme fonctionnait, de tester ses qualités et ses défauts, et j'avoue que les premiers résultats furent plutôt déroutants.

En effet, après avoir consulté plusieurs protocoles de tests pour évaluer le codage MPEG Surround sur divers stimuli, la méthodologie MUSHRA revenait souvent. Elle m'est apparue comme le meilleur moyen pour évaluer un codage MPEG Surround particulier par rapport à un signal 5.1 PCM non compressé, mais surtout c'était le seul moyen de pouvoir comparer directement trois codecs différents, et de hiérarchiser leurs défauts. On pouvait s'attendre, vus les débits testés, à ce que les candidats déterminent assez facilement la référence cachée, et que les codecs obtiennent des moyennes croissantes avec le débit. Une fois la première série de tests réalisée, tout devint beaucoup moins évident : de très nombreuses erreurs sur la référence cachée, une probable erreur de ne pas avoir glissé de signal d'ancre dans les signaux à noter, des différences moins marquées que celles auxquelles on aurait pu s'attendre, probablement lissées par le choix des extraits. Néanmoins, avoir choisi de tester de véritables programmes télévisés était une décision tout à fait réfléchie, puisque l'essentiel de mon mémoire consistait à déterminer si ce codage était adapté à une diffusion télévisuelle, c'est-à-dire par rapport aux types de programmes et aux contraintes de diffusion. La présence d'ancre parmi les signaux à évaluer aurait peut-être permis de simplifier la post-sélection pour l'analyse statistique. De plus, pour ces tests, j'ai fait essentiellement appel à des étudiants de l'école, et à des professionnels du son à la télévision (assistants plateaux, ingénieurs du son, mixeurs télé, ingénierie audio), considérant ces catégories de personnes comme des experts audio. Au vu des résultats, je m'aperçois que ces personnes n'ont peut-être pas

l'habitude d'écouter le son de façon aussi critique, et de comparer divers codages, à l'exception de quelques uns d'entre eux qui ont une oreille très critique et une habitude de ce genre de tests d'écoute.

Mes résultats statistiques ne sont pas suffisamment fiables pour être exploités en tant que tels, toutefois des tendances se dégagent : si le codage MPEG Surround associé au codage HE-AAC à 96 kbits/seconde semble peut différencier de celui à 64 kbits/seconde ou de celui à 128 kbits/seconde, on peut tout de même noter que les deux premiers ont généré beaucoup d'artefacts au niveau des applaudissements, du timbre dans le haut du spectre, et de la largeur stéréophonique. Cependant, il faut tout de même noter que les artefacts produits par un débit de 64 kbits/seconde en MPEG Surround sont dérisoires, en comparaison d'un codage HE-AAC seul, par exemple, au même débit. Il faut aussi tenir compte du fait que les tests ont été réalisés dans un laboratoire, sur un système 5.1 relativement neutre, dans une pièce traitée acoustiquement : il y a fort à parier que sur un système home-cinéma dans un salon quelconque, ces artefacts seraient encore moins perceptibles.

Le codage MPEG Surround associé au codage HE-AAC à 128 kbits/seconde semble le meilleur, les artefacts produits sont acceptables et ce débit est tout à fait envisageable pour des vecteurs limités en bande passante, tels que la télévision numérique en définition standard, la télévision de rattrapage ou encore la vidéo à la demande, et l'on peut même imaginer que s'il était utilisé aussi en diffusion sur la TNT en haute définition, on pourrait envisager d'avoir en plus de la version française en 5.1, la version originale et l'audio-description en 5.1 aussi.

De plus, la deuxième série de tests, bien que réalisée avec très peu de participants, permet de montrer que le downmix stéréophonique créé par un encodeur MPEG Surround est de très bonne qualité : la compatibilité stéréophonique est donc vérifiée au moins du point de vue qualitatif.

La compatibilité en binaural n'a pas pu être testée comme elle est décrite dans la norme, puisque l'encodeur ne proposait qu'un choix réduit d'encodage en binaural.

En revanche, la seule implémentation de l'algorithme d'encodage en MPEG Surround a pour l'instant un énorme inconvénient : le downmix stéréophonique a un niveau de 2 à 3 dB en dessous du niveau du mixage 5.1. Alors que la recommandation R128 est en vigueur pour toutes les chaînes de télévision diffusées en France, il paraît impensable de proposer un format dont les niveaux en stéréophonie et en multicanal ne seraient pas identiques. Néanmoins, puisque le format HE-AAC supporte des métadonnées, et notamment une métadonnée équivalente au Dialnorm de Dolby, on pourrait envisager une évolution du codage MPEG Surround, incluant cette métadonnée dans le flux MPEG-4, permettant au moment du décodage du downmix de l'adapter au bon niveau sonore.

Ce codage apparaît somme toute très étonnant, et ce à tous points de vue. Malgré les artefacts audibles et des résultats peu différenciés, les candidats ont été majoritairement impressionnés par la qualité de l'encodage.

Évidemment, il faudrait mener d'autres tests plus approfondis, avec d'autres extraits de programmes télévisés (par exemple un match de foot mixé en 5.1, une pièce de théâtre, un concert de musique rock très compressé, un concert de musique jazz, etc.), en changeant de protocole, afin de confirmer ou d'infirmer les tendances obtenues. Si ces tests s'avéraient concluants, il pourrait être intéressant d'envisager des essais de diffusion en MPEG Surround, notamment en vidéo à la demande. Néanmoins, le principal frein au développement de ce codage est la prépondérance des systèmes Dolby, inclus dans tous les décodeurs existants, alors que les décodeurs compatibles MPEG Surround ne sont pas encore bien développés.

En outre, un des autres enjeux de ce codage est sa compatibilité binaurale : il pourrait être intéressant de mener une série de tests avec des extraits de programmes regardés sur une tablette ou un smartphone, avec un son binaural, et ce au moyen d'une application qui intégrerait un décodeur MPEG Surround et un encodeur binaural, auquel on pourrait injecter nos propres HRTF. Si les candidats démontrent qu'ils sont conquis

par cette nouvelle technologie, il pourrait alors être intéressant de la développer. Seul bémol : pour l'instant, le format MPEG Surround est soumis à une licence payante, ce qui limite son intégration dans les équipements grand public.

Toutefois, bien que ce codage semble prometteur, il va être difficile de l'implanter dans les mois à venir dans la diffusion de la télévision numérique terrestre car pour le moment, aucun récepteur télévisé ne contient le décodeur adéquat. Une mise à jour ou un changement de matériels seraient les seules solutions pour décoder ce type de flux mais la licence MPEG Surround est payante et ceci engendrerait des coûts supplémentaires. Et si peu de temps après l'avènement de la TNT et l'obligation pour tous les clients d'investir dans de nouveaux décodeurs, ce n'est pas donc pas envisageable à l'heure actuelle.

De plus, il faut rappeler que Dolby est très implanté sur le marché aujourd'hui et qu'il paraît compliqué d'imposer un nouveau format, alors que le Dolby Digital Plus convient à tous : la qualité est plutôt bonne, le flux intègre une métadonnée de niveau des dialogues (Dialnorm), du choix de compression (DRC), ou encore du choix d'écouter un downmix stéréophonique ou un signal multicanal. Ce format est à ce jour le plus satisfaisant et risque de rester implanté encore très longtemps !

En revanche, il paraît beaucoup plus facile à mettre en œuvre le codage MPEG Surround pour des programmes en replay ou en vidéo à la demande, qui pourraient être visionnés sur une tablette, un smartphone ou un ordinateur, avec un petit plugin qui inclurait un décodeur MPEG Surround et un encodage binaural.

Il serait donc intéressant de pouvoir poursuivre des recherches dans ce sens, ainsi que de réaliser d'autres tests perceptifs pour confirmer les premières tendances obtenues : à savoir une faible préférence pour le codec HE-AAC associé au MPEG Surround à 128 kbits/seconde. Si ce format venait à se démocratiser, il faudrait envisager un outil pour que les mixeurs en régie puissent comparer la version PCM et la version encodée, puis décodée en MPEG Surround, afin d'adapter éventuellement leur mixage (sous réserve qu'ils aient le temps !!!).

BIBLIOGRAPHIE

OUVRAGES

BENOIT, Hervé, *La télévision numérique*, Paris, Dunod, coll. Audio-Photo-Vidéo, 1996, 5ème édition 2010.

BLAUERT, Jens, *Spatial Hearing - The Psychophysics of Human Sound Localization*, Traduit par John S. Allen. Edition originale Räumliches Hören, de S. Hirzel Verlag, Stuttgart, 1974. Edition révisée, MIT Press, Cambridge Massachusetts, 1996.

BREEBAART, Jeroen, FALLER, Christof, *Spatial Audio Processing - MPEG Surround and other applications,* Chichester, England, John Wiley & Sons, Octobre 2007.

MADISETTI, Vijay K., *The Digital Signal Processing Handbook*, Boca Raton, FL, CRC Press - Taylor & Francis Group, 2009, réédition CRC Press 2010.

RUMSEY, Francis, *Spatial Audio*, Focal Press/Elsevier, coll. Music Technology Series, United Kingdom, Août 2001, rééd. Focal Press 2005.

PARUTIONS AU JOURNAL DE L'AES

BREEBAART, Jeroen, HOTHO, Gerard, KOPPENS, Jeroen, SCHUIJERS, Erik, OOMEN, Werner, and VAN DE PAR Steven, «Background, Concept, and Architecture for the Recent MPEG Surround Standard on Multichannel Audio Compression », *Journal of the Audio Engineering Society*, Volume 55, n°5, Mai 2007, p. 331-351.

HERRE, Jürgen, KJÖRLING, Kristofer, BREEBAART, Jeroen, FALLER, Christof, DISCH, Sascha, PURNHAGEN, Heiko, KOPPENS, Jeroen, HILPERT, Johannes, RÖDEN, Jonas, OOMEN, Werner, LINZMEIER, Karsten, and CHONG, Kok Seng, «MPEG Surround – The ISO/MPEG standard for efficient and compatible multichannel audio coding», *Journal of the Audio Engineering Society*, Volume 56, n°11, Novembre 2008, p. 932-955.

ARTICLES ISSUS DES CONFÉRENCES DE L'AES

BEACK, Seungkwon, SEO, JEONGIL, JANG, Inseon, JANG, Dae-Young, « Multichannel sound scene for MPEG Surround », *Proceedings of the 29th AES International Conference*, Seoul, KOREA, 2-4 Septembre 2006.

BREEBAART, Jeroen, HERRE, Jürgen, FALLER, Cristof, RÖDEN, Jonas, MYBURG, F., DISCH, Sascha, PURNHAGEN, Heiko, HOTHO, Gerard, NEUSINGER, M., KJÖRLING, Kristofer, and OOMEN, Werner, « MPEG Spatial Audio Coding / MPEG Surround : Overview and Current Status», *Proceedings of the 119th AES Convention,* New-York, USA, 7-10 Octobre 2005.

BREEBAART, Jeroen, HERRE, Jürgen, VILLEMOES, Lars, JIN, Craig, KJÖRLING, Kristofer, PLOGSTIES, Jan, and KOPPENS, Jeroen, « Multi-channel goes mobile : MPEG Surround Binaural Rendering», *Proceedings of the 29th AES Conference*, Seoul, SOUTH KOREA, 2-4 Septembre 2006.

DIETZ, Martin, LILJERYD, Lars, KJÖRLING, Kristofer, KUNZ, Oliver, « Spectral Band Replication, a novel approach in audio coding », *Proceedings of the 112th AES Convention*, Munich, GERMANY, 10-13 Mai 2002.

HERRE, Jürgen, KJÖRLING, Kristofer, BREEBAART, Jeroen, FALLER, Cristof, DISCH, Sascha, PURNHAGEN, Heiko, KOPPENS, Jeroen, HILPERT, Johannes, RÖDEN, Jonas, OOMEN, Werner, LINZMEIER, Karsten, and CHONG, Kok Seng, « MPEG Surround – The ISO/MPEG Standard for Efficient and Compatible Multi-Channel Audio Coding», *Proceedings of the 122nd AES Convention,* Vienna, AUSTRIA, 5-8 Mai 2007.

HERRE, Jürgen, PURNHAGEN, Heiko, BREEBAART, Jeroen, FALLER, Cristof, DISCH, Sascha, KJÖRLING, Kristofer, SCHUIJERS, Erik, HILPERT, Johannes, MYBURG, François, « The Reference Model Architecture for MPEG Spatial Audio Coding », *Proceedings of the 118th AES Convention*, Barcelona, SPAIN, 28-31 Mai 2005.

RÖDEN, Jonas, BREEBAART, Jeroen, HILPERT, Johannes, PURNHAGEN, Heiko, SCHUIJERS, Erik, KOPPENS, Jeroen, LINZMEIER, Karsten, and HÖLZER, Andreas, « A study of the MPEG Surround quality versus bit-rate curve», *Proceedings of the 123rd AES Convention*, New-York, USA, 5-8 Octobre 2007.

NORMES ET RECOMMANDATIONS

Norme ISO/IEC 23003-1:2007 – *Information Technology – MPEG Audio Technologies* – « Part 1 : MPEG Surround », 2007.

Recommandation UIT-R BS.1116-1 - *Méthodes d'évaluation subjective des dégradations faibles dans les systèmes audio, y compris les systèmes sonores multi-voies,* 1994-1997.

Recommandation UIT-R BS.1534 - *Méthode d'évaluation subjective du niveau de qualité intermédiaire des systèmes de codage*, 2001.

Documentation technique EBU/UER TECH 3324 - *EBU Evaluations of Multichannel Audio Codecs*, Genève, Septembre 2007.

Recommandation EBU/UER R128 - *Loudness normalisation and permitted maximum level of audio signal*, Genève, Août 2011.

Recommandation ITU-R BS.1770-2 - *Algorithms to measure audio programme loudness and true peak audio level*, Genève, Mars 2011.

Recommandation ITU-R BS.775-3 - *Système de son stéréophonique multicanal avec ou sans image associée*, Genève, Août 2012.

MÉMOIRES, ÉTUDES

POMMERET, Pierre, *Diffusion audio multi-formats à la télévision*, Mémoire sous la direction de Jean Châtauret et de Jean-Yves Carabot, Son pour la télévision, Ecole Nationale Supérieure Louis Lumière, 93, Juin 2007

LIBOLT, Anaïs, *La création des métadonnées du Dolby E en prévision d'une diffusion multicanale en Dolby Digital*, Mémoire sous la direction d'Alain Delhaise et Christian Bourguignon, Ecole Nationale Supérieure Louis Lumière, 93, 23 juin 2006.

DANARD, Benoît, DAUBARD, Sophie, MALHERBE, Clément, *Les usages de la télévision connectée*, Etude du Centre National du Cinéma et de l'image animée, Paris, Décembre 2012.

MERIENNE, Claire, «Codecs : le MPEG Surround», Septembre 2011
URL : http://aesfrance.info/index.php/broadcast/30-codecs-le-mpeg-surround , page consultée le 21/01/2013.

Advanced Television Systems Committee Inc., « Digital Audio Compression Standard (AC-3, E-AC-3) », Document A/52 :2010, 22 Novembre 2010.

SITES INTERNET

Site qui détaille le MPEG Surround, URL : http://www.mpegsurround.com/ , site consulté le 11/03/2012.

FUCHS, Harald, KORTE, Olaf, HILPERT, Johannes, « Digital Broadcasting with MPEG Surround», 2009,
URL : http://tech.ebu.ch/docs/techreview/trev_2009-Q3_MPEG_Fraunhofer.pdf, page consultée le 18/03/2012

Site du fabricant **Fraunhofer** qui propose de nombreux liens vers des documentations sur le codec MPEG Surround. Site consulté le 18/03/2012.
URL :http://www.iis.fraunhofer.de/en/bf/amm/produkte/audiocodec/audiocodecs/mpegsurr/

Page du fabricant **Fraunhofer** qui expose les caractéristiques du plug-in « Fraunhofer Pro-Codec » de Sonnox. Page consultée le 15/04/2012.
URL : http://www.iis.fraunhofer.de/en/bf/amm/produkte/audiocodec/audiocodecs/sonnox-plugin/

Site du fabricant **Sonnox**, Documentation du plug-in « Fraunhofer Pro-Codec » de Sonnox.
URL : http://www.sonnoxplugins.com/pub/plugins/products/pro-codec.htm , site consulté le 31/03/2012.

Site qui compare les débits audio et vidéo des chaînes de télévision européennes entre les différents systèmes de réception (TNT, FAI, câble, etc.). URL : http://www.digitalbitrate.com/ , site consulté le 01/04/2013

Site de **Jeroen BREEBAART**, membre de l'IEEE et de l'AES, a travaillé pour le laboratoire Philips Research, actuellement chez Dolby Laboratories in Sydney. Site qui contient tous les articles qu'il a publié. URL : http://www.jeroenbreebaart.com/ , page consultée le 21/01/2013

Site du **Conseil Supérieur de l'Audiovisuel**, qui contient des études sur l'évolution de la télévision. URL : http://www.csa.fr/ . Page consultée le 12/02/2013.

Site du **SIMAVELEC**, le syndicat des industries de matériels audiovisuels électroniques. URL : http://www.simavelec.net/ . Page consultée le 12/02/2013.

Site de **Fraunhofer**, White paper « HE-AAC Metadata for digital broadcasting », Septembre 2011.
URL : http://www.iis.fraunhofer.de/content/dam/iis/en/dokumente/AMM/Metadata-whitepaper.pdf , page consultée le 16/04/2013.

Site de **Mediamétrie**, institut de mesures de l'audimat des chaînes de télévision et des stations radio. URL : http://www.mediametrie.fr , page consultée le 17/04/2013.

RAKOTOMALALA, Ricco, « Tests de normalité, Techniques empiriques et tests statistiques », Version 2.0, Université Lumière Lyon 2, 1er octobre 2011.

URL : http://eric.univ-lyon2.fr/~ricco/cours/cours/Test_Normalite.pdf , page consultée le 06/05/2013

EZRATTY, Olivier, Opinions Libres – Le blog d'Olivier Ezratty, Innovation et médias numériques, « Les infrastructures de France Télévisions – Workflow numérique », article publié le 12 février 2012 et mis à jour le 14 février 2012.
URL : http://www.oezratty.net/wordpress/2012/les-infrastructures-de-france-televisions-workflow-numerique/ , page consultée le 17/05/2013.

Site de **TDF**, TéléDiffusion de France, concepteur et diffuseur historique français de réseaux télécoms (TNT, radio, téléphonie mobile, etc.).
URL : http://www.tdf.fr , page consultée le 16/05/2013.

TABLE DES ILLUSTRATIONS

ANNEXES

ANNEXE A : COMPARAISON DES FLUX AUDIO ET VIDÉO DES CHAÎNES TÉLÉVISÉES EN FONCTION DES VECTEURS DE DIFFUSION

Relevé réalisé sur plusieurs chaînes le 1er avril 2013 sur le site internet http://www.digitalbitrate.com/. Ici le relevé de France 3 et France 2 (SD et HD).

	France 3		
	Image	Son	Data
TNT SD (R1)	MPEG2 Résolution 544*576 Format 1,77 (16:9) Débit fixe 4,4 Mbits/s	MPEG2 joint stereo à 48kHz Paire 1 192 kbits/s Paire 2 192 kbits/s	Sous-titres 1/5/17 kbits/s 0x6f 1 kbits/s 0x13 1/2/2 kbits/s
TNT HD	Pas de diffusion HD sur France 3	Pas de diffusion HD sur France 3	Pas de diffusion HD sur France 3
FAI FREE ADSL Bas débit	MPEG 4 Résolution 720*576 Format 16:9 Débit fixe : 1,4 Mbits/s	MPEG 4 48 kHz 112 kbits/s	
FAI FREE ADSL SD	MPEG4 Résolution 720*576 Format ? Débit min : 1,4 Mbits/s Débit moy : 1,5 Mbits/s Débit max : 1,7 Mbits/s	MPEG2 Joint 48 kHz Paire 1 192 kbits/s	
FAI FREE ADSL HD	MPEG4 Résolution 1440*1080 Format 16:9 Débit min : 4,1 Mbits/s Débit moy : 4,4 Mbits/s Débit max : 4,6 Mbits/s	MPEG 4 à 48 kHz Paire 1 ? Paire 2 : AC3 2.0 à 128 ou 160 kbits/s	
Câble SD (Numericable)	MPEG 2 Résolution 544*576 Format 16:9 Débit min : 2,1 Mbits/s Débit moy : 4 Mbits/s Débit max : 8,2 Mbits/s	MPEG 2 joint 48 kHz Paire 1 : joint à 192 kbits/s	Sous-titres 2/9/26 kbits/s
EUTELSAT 5°W (ch	France 3 (11591 V) Résolution 544*576 Format 16:9 Débit min : 4,4 Mbits/s Débit moy : 4,5 Mbits/s Débit max : 4,5 Mbits/s	MPEG 2 Joint à 48 kHz Paire 1 MPEG joint à 192 kbits/s Paire 2 MPEG joint à 192 kbits/s	Sous-titres 2/8/25 kbits/s Télétexte 26/26/27 kbits/s Privé 1/2/2 kbits/s Oxb 2/3/3 kbits/s

	France 2 et France 2 HD		
	Image	Son	Data
TNT SD (R1)	MPEG2 Résolution 720*576 Format 1,77 (16:9) Débit min : 2 Mbits/s Débit moy : 3,8 Mbits/s Débit max : 8,1 Mbits/s	MPEG2 joint stereo à 48kHz Paire 1 192 kbits/s Paire 2 192 kbits/s Paire 3 192 kbits/s	Sous-titres 1/4/16 kbits/s Sous-titres 1/1/2 kbits/s 0x6f 1 kbits/s
TNT HD (R5)	MPEG4 Résolution 1920*1080 Format 1,77 (16:9) Débit min : 4 Mbits/s Débit moy : 6,7 Mbits/s Débit max : 13 Mbits/s	E-AC3 à 48 kHz Paire 1 : 5.1 à 256 kbits/s Paire 2 : AC3 2.0 à 128 kbits/s Paire 3 : AC3 2.0 à 128 kbits/s	Sous-titres 1/4/13 kbits/s Sous-titres 1/1/2 kbits/s 0x6f 1 kbits/s 0x13 48 kbits/s
FAI FREE ADSL Bas débit	MPEG4 Résolution 720*576 Format 16:9 Débit min : 1,4 Mbits/s Débit moy : 1,5 Mbits/s Débit max : 1,7 Mbits/s	MPEG 4 48 kHz Paire 1 ? Paire 2 AC3 2.0 à 128 ou 192 kbits/s	
FAI FREE ADSL SD	MPEG4 Résolution 720*576 Format ? Débit min : 0,8 Mbits/s Débit moy : 1,1 Mbits/s Débit max : 1,4 Mbits/s	MPEG 2 Joint 48 kHz Paire 1 192 kbits/s	
FAI FREE ADSL HD	MPEG 4 Résolution 1440*1080 Format 16:9 Débit min : 3,8 Mbits/s Débit moy : 4 Mbits/s Débit max : 4,4 Mbits/s	MPEG 4 à 48 kHz Paire 1 ? Paire 2 : AC3 2.0 à 128 kbits/s	
Câble SD (Numericable)	MPEG 2 Résolution 720*576 Format 16:9 Débit min : 2 Mbits/s Débit moy : 4,2 Mbits/s Débit max : 8,2 Mbits/s	MPEG 2 joint 48 kHz Paire 1 : joint à 192 kbits/s Paire 2 : joint à 192 kbits/s	Sous-titres 2/5/20 kbits/s
Câble HD (Numericable)	MPEG 4 Résolution 1920*1080 Format 1,77 (16:9) Débit min : 4,1 Mbits/s Débit moy : 6,8 Mbits/s Débit max : 13,2 Mbits/s	E-AC3 à 48 kHz Paire 1 : e-AC3 2.0 à 128 kbits/s Paire 2 : e-AC3 5.1 à 256 kbits/s	Sous-titres 2/6/21 kbits/s
Satellite EUTELSAT 5°W France 2 SD (11591V)	MPEG2 Résolution 720*576 Format 16:9 Débit min : 2 Mbits/s Débit moy : 4,2 Mbits/s Débit max : 8,2 Mbits/s	MPEG 2 joint à 48 kHz Paire 1 MPEG joint à 192 kbits/s Paire 2 MPEG joint à 192 kbits/s Paire 3 MPEG joint à 192 kbits/s	Sous-titres 2/8/28 kbits/s Sous-titres 2/2/3 kbits/s Télétexte 14 kbits/s Privé 1/1/2 kbits/s
Satellite EUTELSAT 5°W France 2 HD (11096V)	MPEG 4 Résolution 1920*1080 Format 16:9 et autre Débit min : 4,1 Mbits/s Débit moy : 6,7 Mbits/s Débit max : 13,3 Mbits/s	E-AC3 2.0 ou 5.1 à 48 kHz ou crypté Paire 1 : E-AC3 5.1 à 256 kbits/s Paire 2 : E-AC3 2.0 à 128 kbits/s	Sous-titres 2/8/28 kbits/s Privé 1/2/2 kbits/s Oxb 94/94/95 kbits/s Sous-titres 2/2/3 kbits/s

ANNEXE B : 1ER TEST PERCEPTIF

CONSIGNES POUR LE TEST PERCEPTIF :

Session N°
Date

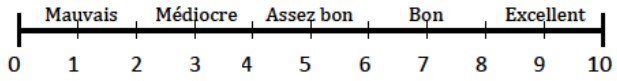
Tests perceptifs :
Evaluer le codage MPEG Surround sur différents extraits de programmes télévisuels.

Bonjour et bienvenue au laboratoire Innovations&Développement de France Télévisions.

Vous allez participer à une séance de tests perceptifs afin d'évaluer les qualités et les défauts du MPEG Surround. Pour chaque extrait image et son d'une durée d'environ 30 secondes, vous allez d'abord écouter le programme original : il s'agit de la référence. Ensuite, vous écouterez quatre stimuli, nommés A, B, C et D, parmi lesquels une référence cachée, c'est-à-dire le signal original, s'est glissée, ainsi que trois stimuli encodés et décodés en MPEG Surround, ces quatre signaux étant répartis dans un ordre aléatoire.

Pour chaque extrait, veuillez donner votre appréciation globale du signal audio pour chaque stimulus. Pour noter les extraits, vous placerez une croix sur une droite graduée de 0 à 10 : la note 0 signifie un signal très dégradé, de très mauvaise qualité, totalement inacceptable, tandis que la note 10 définit un signal excellent, de même qualité que l'original. Pour chaque extrait, l'un des stimuli est forcément identique à l'original, et doit obtenir la note de 10. Si deux stimuli vous paraissent de même qualité que l'original, ils peuvent obtenir tous deux la note de 10.

Voici un exemple de droite graduée utilisée pour les notations, avec des termes d'évaluation qui peuvent vous guider.



Vous pouvez gérer le transport depuis l'interface Command 8 et choisir le stimulus écouté par l'intermédiaire des « solo » (le ProTools est paramétré en mode solo X-OR : quand on appuie sur un bouton « solo » d'une piste, cette action annule le précédent « solo »). L'extrait numéro 1 débute au time-code 00 :01 :00 :00, l'extrait numéro 2 débute à 00 :02 :00 :00, et ainsi de suite. Pour passer à l'extrait suivant, vous pouvez utiliser la souris pour sélectionner le marqueur correspondant dans la fenêtre « emplacement mémoire ».

Vous êtes libres d'écouter les stimuli en entier, les uns après les autres, ou de zapper de l'un à l'autre, à condition de ne pas dépasser le temps imposé : à savoir 5 minutes par extrait.

Le test peut commencer dès que vous êtes prêts.

NOTES DE TOUS LES CANDIDATS, POUR CHAQUE EXTRAIT

Extrait	Participant	Référence cachée	MPS 64k	MPS 96k	MPS 128k
Extrait 1	1	10,00	7,00	10,00	10,00
	2	10,00	9,00	9,00	9,00
	3	7,10	5,00	10,00	8,10
	4	9,00	10,00	6,00	9,00
	5	10,00	8,50	8,50	8,00
	6	6,00	6,00	9,70	10,00
	7	10,00	8,00	6,00	9,00
	8	6,00	7,00	5,00	8,00
	9	10,00	6,00	10,00	7,00
	10	7,00	10,00	10,00	7,00
	11	10,00	8,00	9,00	9,00
	12	10,00	2,00	8,00	8,00
	14	10,00	5,00	5,00	10,00
	15	10,00	7,00	8,00	9,00
	16	5,00	2,00	10,00	8,00
	17	9,00	7,00	10,00	6,00
	18	6,00	9,00	9,00	3,00
	19	10,00	3,00	5,00	8,00
	20	8,00	10,00	6,00	9,00
	21	6,90	10,00	8,30	7,50
	22	10,00	8,50	10,00	8,50
	23	5,40	10,00	4,50	3,50
	24	9,00	10,00	9,00	9,00
	25	8,00	2,00	4,00	10,00
	26	10,00	5,00	6,00	9,00
	27	9,00	9,00	8,00	9,00
	28	10,00	3,00	5,00	9,00
	29	10,00	9,00	9,00	10,00
MOYENNE EXT 1		8,62	7,00	7,79	8,24
ECART-TYPE EXT 1		1,721	2,705	2,066	1,724

Extrait	Participant	Référence cachée	MPS 64k	MPS 96k	MPS 128k
Extrait 2	1	10,00	7,00	10,00	10,00
	2	10,00	9,00	9,00	9,00
	3	10,00	7,00	6,00	5,00
	4	10,00	6,00	10,00	8,00
	5	4,00	10,00	8,00	4,00
	6	10,00	4,90	2,10	7,80
	7	10,00	9,00	4,00	8,00
	8	7,00	6,00	8,00	6,00
	9	10,00	6,00	6,00	9,00
	10	10,00	10,00	8,00	10,00
	11	10,00	6,00	7,00	5,00
	12	6,80	4,90	2,80	10,00
	14	10,00	5,00	5,00	7,00
	15	10,00	7,00	8,00	5,00
	16	10,00	8,00	2,00	5,00
	17	10,00	7,50	5,00	4,70
	18	8,00	8,00	9,00	6,00
	19	7,50	3,00	10,00	4,50
	20	10,00	8,00	8,00	5,00
	21	10,00	9,60	9,60	9,60
	22	10,00	5,00	6,00	7,00
	23	10,00	10,00	10,00	10,00
	24	9,00	9,00	9,00	10,00
	25	10,00	5,00	5,00	7,00
	26	10,00	6,00	7,00	7,00
	27	8,00	9,00	9,00	9,00
	28	10,00	7,00	6,00	9,00
	29	10,00	8,00	8,00	8,00
MOYENNE EXT 2		9,30	7,18	7,05	7,34
ECART-TYPE EXT 2		1,448	1,875	2,406	2,024

Extrait	Participant	Référence cachée	MPS 64k	MPS 96k	MPS 128k
Extrait 3	1	7,00	10,00	10,00	10,00
	2	10,00	10,00	10,00	10,00
	3	6,00	7,00	7,00	10,00
	4	10,00	8,00	10,00	10,00
	5	3,00	6,00		10,00
	6	5,00	5,10	9,70	4,40
	7	10,00	7,00	7,00	9,00
	8	8,00	7,00	8,00	7,00
	9	6,00	9,00	6,00	10,00
	10	10,00	10,00	10,00	10,00
	11	9,00	9,00	9,00	9,00
	12	4,90	7,90	8,10	10,00
	14	10,00	10,00	10,00	10,00
	15	10,00	10,00	8,00	6,00
	16	5,00	2,00	8,00	10,00
	17	10,00	5,00	5,00	7,00
	18	3,00	3,00	6,00	7,00
	19	10,00	3,00	4,00	7,00
	20	9,40	9,30	10,00	9,30
	21	10,00	9,40	8,50	10,00
	22	10,00	10,00	9,50	9,50
	23	10,00	10,00	10,00	10,00
	24	9,00	9,00	9,00	10,00
	25	10,00	8,00	5,00	8,00
	26	9,00	9,00	10,00	9,00
	27	9,00	9,00	9,00	9,00
	28	3,00	3,00	3,00	3,00
	29	9,00	9,00	10,00	10,00
MOYENNE EXT 3		8,05	7,67	8,14	8,72
ECART-TYPE EXT 3		2,483	2,509	2,083	1,879

Extrait	Participant	Référence cachée	MPS 64k	MPS 96k	MPS 128k
Extrait 4	1	10,00	7,00	10,00	10,00
	2	10,00	10,00	10,00	10,00
	3	8,00	5,00	7,00	10,00
	4	8,00	8,00	8,00	10,00
	5	10,00	5,00	7,00	8,00
	6	9,40	6,20	9,40	8,40
	7	7,00	9,00	9,00	10,00
	8	7,00	8,00	8,00	5,00
	9	10,00	10,00	9,00	6,00
	10	10,00	10,00	10,00	10,00
	11	10,00	6,00	7,00	7,00
	12	8,00	10,00	5,00	3,00
	14	10,00	10,00	10,00	10,00
	15	10,00	7,00	8,00	9,00
	16	10,00	8,00	2,00	2,00
	17	7,00	10,00	7,70	5,30
	18	4,00	8,00	10,00	6,00
	19	3,00	8,00	10,00	6,00
	20	8,20	10,00	8,40	8,30
	21	10,00	7,70	8,20	8,40
	22	10,00	6,00	7,00	10,00
	23	8,80	9,00	7,90	9,00
	24	9,00	9,00	9,00	10,00
	25	10,00	7,00	7,00	7,00
	26	9,00	6,00	6,00	10,00
	27	7,00	6,00	9,00	9,00
	28	6,00	10,00	8,00	6,00
	29	10,00	9,00	10,00	9,00
MOYENNE EXT 4		8,55	8,03	8,13	7,94
ECART-TYPE EXT 4		1,897	1,681	1,803	2,274

Extrait	Participant	Référence cachée	MPS 64k	MPS 96k	MPS 128k
Extrait 5	1	8,00	10,00	10,00	10,00
	2	10,00	6,00	6,00	5,00
	3	10,00	8,00	4,00	2,00
	4	10,00	6,00	7,00	6,00
	5	10,00	8,00	4,50	3,50
	6	9,70	1,10	9,10	8,10
	7	9,00	7,00	4,00	10,00
	8	8,00	7,00	6,00	7,00
	9	10,00	6,00	8,00	5,00
	10	10,00	10,00	9,00	10,00
	11	10,00	6,00	8,00	9,00
	12	10,00	4,00	8,00	2,00
	14	10,00	9,00	7,00	7,00
	15	10,00	6,00	8,00	9,00
	16	2,00	5,00	8,00	10,00
	17	10,00	5,00	7,30	4,70
	18	8,00	3,00	8,00	4,00
	19	3,00	5,00	7,00	10,00
	20	8,00	8,00	10,00	4,00
	21	9,20	9,20	9,50	10,00
	22	10,00	8,00	10,00	9,00
	23	9,70	9,60	9,80	9,80
	24	9,00	10,00	9,00	9,00
	25	10,00	2,00	7,00	6,00
	26	10,00	6,00	5,00	8,00
	27	10,00	8,00	6,00	8,00
	28	6,00	6,00	8,50	10,00
	29	10,00	10,00	8,00	10,00
MOYENNE EXT 5		8,91	6,75	7,56	7,36
ECART-TYPE EXT 5		2,062	2,424	1,777	2,659

Extrait	Participant	Référence cachée	MPS 64k	MPS 96k	MPS 128k
Extrait 6	1	10,00	7,00	10,00	9,00
	2	10,00	10,00	10,00	10,00
	3	9,10	9,00	9,10	10,00
	4	10,00	10,00	10,00	9,00
	5	10,00	5,00	6,00	8,00
	6	4,10	10,00	8,90	6,50
	7	7,50	9,00	10,00	9,00
	8	7,00	7,00	8,00	7,00
	9	9,00	10,00	7,00	9,00
	10	10,00	10,00	10,00	10,00
	11	9,00	9,00	8,00	9,00
	12	8,00	8,00	7,90	8,00
	14	9,00	7,00	10,00	9,00
	15	8,00	7,00	9,00	10,00
	16	8,00	2,00	10,00	5,00
	17	5,00	7,30	10,00	5,00
	18	8,80	10,00	8,00	5,90
	19	5,00	3,00	10,00	7,50
	20	10,00	9,30	9,30	9,30
	21	8,80	8,50	10,00	9,00
	22	6,00	7,50	9,00	10,00
	23	9,70	9,40	9,30	9,50
	24	9,00	9,00	10,00	9,00
	25	10,00	7,00	4,00	7,00
	26	9,00	7,00	10,00	9,00
	27	9,00	8,00	9,00	8,00
	28	10,00	8,00	10,00	6,50
	29	9,00	10,00	9,00	9,00
MOYENNE EXT 6		8,50	8,00	8,98	8,33
ECART-TYPE EXT 6		1,676	2,043	1,432	1,492

Extrait	Participant	Référence cachée	MPS 64k	MPS 96k	MPS 128k
Extrait 7	1	10,00	10,00	7,00	10,00
	2	9,00	9,00	9,00	10,00
	3	10,00	8,20	7,00	9,10
	4	8,00	7,00	7,00	10,00
	5	10,00	3,50	9,00	7,00
	6	9,40	3,30	2,80	6,90
	7	9,00	10,00	8,00	9,00
	8	8,00	8,00	8,00	8,00
	9	10,00	8,00	8,00	9,00
	10	10,00	8,00	10,00	10,00
	11	10,00	4,00	6,00	8,00
	12	10,00	3,90	6,90	8,00
	14	10,00	7,00	10,00	10,00
	15	10,00	8,00	6,00	7,00
	16	2,00	5,00	10,00	8,00
	17	10,00	8,60	8,60	8,40
	18	8,00	6,00	9,00	5,00
	19	5,00	3,50	8,00	10,00
	20	6,00	5,00	10,00	7,00
	21	9,50	10,00	9,20	9,20
	22	10,00	10,00	9,00	9,00
	23	9,00	10,00	9,00	10,00
	24	10,00	8,00	8,00	8,00
	25	10,00	2,00	5,00	5,00
	26	10,00	4,00	8,00	8,00
	27	9,00	8,00	8,00	8,00
	28	8,00	5,00	6,00	10,00
	29	10,00	9,00	10,00	9,00
MOYENNE EXT 7		8,93	6,86	7,95	8,45
ECART-TYPE EXT 7		1,858	2,481	1,710	1,432

MOYENNES DES CODECS SUR L'ENSEMBLE DES EXTRAITS

POUR CHAQUE PARTICIPANT

	Participant	Référence cachée	MPS 64k	MPS 96k	MPS 128k
Moyenne	1	9,29	8,29	9,57	9,86
Ecart-type P		1,161	1,485	1,050	0,350
Moyenne	2	9,86	9,00	9,00	9,00
Ecart-type P		0,350	1,309	1,309	1,690
Moyenne	3	8,60	7,03	7,16	7,74
Ecart-type P		1,488	1,436	1,822	2,872
Moyenne	4	9,29	7,86	8,29	8,86
Ecart-type P		0,881	1,552	1,578	1,355
Moyenne	5	8,14	6,57	7,17	6,93
Ecart-type P		2,949	2,145	1,546	2,178
Moyenne	6	7,66	5,23	7,39	7,44
Ecart-type P		2,336	2,541	3,139	1,624
Moyenne	7	8,93	8,43	6,86	9,14
Ecart-type P		1,147	1,050	2,167	0,639
Moyenne	8	7,29	7,14	7,29	6,86
Ecart-type P		0,700	0,639	1,161	0,990
Moyenne	9	9,29	7,86	7,71	7,86
Ecart-type P		1,385	1,726	1,385	1,726
Moyenne	10	9,57	9,71	9,57	9,57
Ecart-type P		1,050	0,700	0,728	1,050
Moyenne	11	9,71	6,86	7,71	8,00
Ecart-type P		0,452	1,726	1,030	1,414
Moyenne	12	8,24	5,81	6,67	7,00
Ecart-type P		1,798	2,645	1,887	2,976
Moyenne	14	9,86	7,57	8,14	9,00
Ecart-type P		0,350	1,990	2,231	1,309
Moyenne	15	9,71	7,43	7,86	7,86
Ecart-type P		0,700	1,178	0,833	1,726
Moyenne	16	6,00	4,57	7,14	6,86
Ecart-type P		3,162	2,499	3,356	2,748
Moyenne	17	8,71	7,20	7,66	5,87
Ecart-type P		1,829	1,674	1,932	1,283
Moyenne	18	6,54	6,71	8,43	5,27
Ecart-type P		2,097	2,603	1,178	1,270
Moyenne	19	6,21	4,07	7,71	7,57
Ecart-type P		2,776	1,741	2,312	1,860
Moyenne	20	8,51	8,51	8,81	7,41
Ecart-type P		1,317	1,627	1,374	2,002

	Participant	Référence cachée	MPS 64k	MPS 96k	MPS 128k
Moyenne Ecart-type P	21	9,20 1,032	9,20 0,776	9,04 0,657	9,10 0,840
Moyenne Ecart-type P	22	9,43 1,400	7,86 1,747	8,64 1,432	9,00 0,964
Moyenne Ecart-type P	23	8,94 1,510	9,71 0,368	8,64 1,824	8,83 2,202
Moyenne Ecart-type P	24	9,14 0,350	9,14 0,639	9,00 0,535	9,29 0,700
Moyenne Ecart-type P	25	9,71 0,700	4,71 2,491	5,29 1,161	7,14 1,457
Moyenne Ecart-type P	26	9,57 0,495	6,14 1,457	7,43 1,841	8,57 0,904
Moyenne Ecart-type P	27	8,71 0,881	8,14 0,990	8,29 1,030	8,57 0,495
Moyenne Ecart-type P	28	7,57 2,499	6,00 2,390	6,64 2,183	7,64 2,401
Moyenne Ecart-type P	29	9,71 0,452	9,14 0,639	9,14 0,833	9,29 0,700

TEST : CALCULS DES DIFFÉRENCES LA NOTE ATTRIBUÉE À CHAQUE CODEC ET CELLE ATTRIBUÉE À LA RÉFÉRENCE CACHÉE

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Ext1	MPS 64k	-3,00	-1,00	-2,10	1,00	-1,50	0,00	-2,00	1,00	-4,00	3,00
	MPS 96k	0,00	-1,00	2,90	-3,00	-1,50	3,70	-4,00	-1,00	0,00	3,00
	MPS 128k	0,00	-1,00	1,00	0,00	-2,00	4,00	-1,00	2,00	-3,00	0,00
Ext2	MPS 64k	-3,00	-1,00	-3,00	-4,00	6,00	-5,10	-1,00	-1,00	-4,00	0,00
	MPS 96k	0,00	-1,00	-4,00	0,00	4,00	-7,90	-6,00	1,00	-4,00	-2,00
	MPS 128k	0,00	-1,00	-5,00	-2,00		-2,20	-2,00	-1,00	-1,00	0,00
Ext3	MPS 64k	3,00	0,00	1,00	-2,00	3,00	0,10	-3,00	-1,00	3,00	0,00
	MPS 96k	3,00	0,00	1,00	0,00	0,00	4,70	-3,00	0,00	0,00	0,00
	MPS 128k	3,00	0,00	4,00	0,00	7,00	-0,60	-1,00	-1,00	4,00	0,00
Ext4	MPS 64k	-3,00	0,00	-3,00	0,00	-5,00	-3,20	2,00	1,00	0,00	0,00
	MPS 96k	0,00	0,00	-1,00	0,00	-3,00	0,00	2,00	1,00	-1,00	0,00
	MPS 128k	0,00	0,00	2,00	2,00	-2,00	-1,00	3,00	-2,00	-4,00	0,00
Ext5	MPS 64k	2,00	-4,00	-2,00	-4,00	-2,00	-8,60	-2,00	-1,00	-4,00	0,00
	MPS 96k	2,00	-4,00	-6,00	-3,00	-5,50	-0,60	-5,00	-2,00	-2,00	-1,00
	MPS 128k	2,00	-5,00	-8,00	-4,00	-6,50	-1,60	1,00	-1,00	-5,00	0,00
Ext6	MPS 64k	-3,00	0,00	-0,10	0,00	-5,00	5,90	1,50	0,00	1,00	0,00
	MPS 96k	0,00	0,00	0,00	0,00	-4,00	4,80	2,50	1,00	-2,00	0,00
	MPS 128k	-1,00	0,00	0,90	-1,00	-2,00	2,40	1,50	0,00	0,00	0,00
Ext7	MPS 64k	0,00	0,00	-1,80	-1,00	-6,50	-6,10	1,00	0,00	-2,00	-2,00
	MPS 96k	-3,00	0,00	-3,00	-1,00	-1,00	-6,60	-1,00	0,00	-2,00	0,00
	MPS 128k	0,00	1,00	-0,90	2,00	-3,00	-2,50	0,00	0,00	-1,00	0,00
MOYENNE		-0,048	-0,857	-1,290	-0,952	-1,525	-0,971	-0,786	-0,190	-1,476	0,048
ECART-TYPE		2,061	1,558	2,989	1,830	3,864	4,268	2,528	1,078	2,379	1,161
P (m=0)		0,917	0,020	0,062	0,027	0,094	0,309	0,170	0,428	0,010	0,853
Post-sélection au seuil de 10%		Eliminé	Ok	Ok	Ok	Ok	Eliminé	Eliminé	Eliminé	Ok	Eliminé
Post-sélection au seuil de 25%		Eliminé	Ok	Ok	Ok	Ok	Eliminé	Ok	Eliminé	Ok	Eliminé
2ème post-sélection		Eliminé	Ok	Ok	Ok	éliminé	Eliminé	Ok	Eliminé	Ok	Eliminé

		P11	P12	P14	P15	P16	P17	P18	P19	P20
Ext1	MPS 64k	-2,00	-8,00	-5,00	-3,00	-3,00	-2,00	3,00	-7,00	2,00
	MPS 96k	-1,00	-2,00	-5,00	-2,00	5,00	1,00	3,00	-5,00	-2,00
	MPS 128k	-1,00	-2,00	0,00	-1,00	3,00	-3,00	-3,00	-2,00	1,00
Ext2	MPS 64k	-4,00	-1,90	-5,00	-3,00	-2,00	-2,50	0,00	-4,50	-2,00
	MPS 96k	-3,00	-4,00	-5,00	-2,00	-8,00	-5,00	1,00	2,50	-2,00
	MPS 128k	-5,00	3,20	-3,00	-5,00	-5,00	-5,30	-2,00	-3,00	-5,00
Ext3	MPS 64k	0,00	3,00	0,00	0,00	-3,00	-5,00	0,00	-7,00	-0,10
	MPS 96k	0,00	3,20	0,00	-2,00	3,00	-5,00	3,00	-6,00	0,60
	MPS 128k	0,00	5,10	0,00	-4,00	5,00	-3,00	4,00	-3,00	-0,10
Ext4	MPS 64k	-4,00	2,00	0,00	-3,00	-2,00	3,00	4,00	5,00	1,80
	MPS 96k	-3,00	-3,00	0,00	-2,00	-8,00	0,70	6,00	7,00	0,20
	MPS 128k	-3,00	-5,00	0,00	-1,00	-8,00	-1,70	2,00	3,00	0,10
Ext5	MPS 64k	-4,00	-6,00	-1,00	-4,00	3,00	-5,00	-5,00	2,00	0,00
	MPS 96k	-2,00	-2,00	-3,00	-2,00	6,00	-2,70	0,00	4,00	2,00
	MPS 128k	-1,00	-8,00	-3,00	-1,00	8,00	-5,30	-4,00	7,00	-4,00
Ext6	MPS 64k	0,00	0,00	-2,00	-1,00	-6,00	2,30	1,20	-2,00	-0,70
	MPS 96k	-1,00	-0,10	1,00	1,00	2,00	5,00	-0,80	5,00	-0,70
	MPS 128k	0,00	0,00	0,00	2,00	-3,00	0,00	-2,90	2,50	-0,70
Ext7	MPS 64k	-6,00	-6,10	-3,00	-2,00	3,00	-1,40	-2,00	-1,50	-1,00
	MPS 96k	-4,00	-3,10	0,00	-4,00	8,00	-1,40	1,00	3,00	4,00
	MPS 128k	-2,00	-2,00	0,00	-3,00	6,00	-1,60	-3,00	5,00	1,00
MOYENNE		-2,190	-1,748	-1,619	-2,000	0,190	-1,805	0,262	0,238	-0,267
ECART-TYPE		1,834	3,713	2,085	1,703	5,372	2,945	2,975	4,636	2,045
P (m=0)		<0,0001	0,043	0,002	<0,0001	0,873	0,011	0,691	0,816	0,557
Post-sélection au seuil de 10%		ok	ok	ok	ok	éliminé	ok	éliminé	éliminé	éliminé
Post-sélection au seuil de 25%		ok	ok	ok	ok	éliminé	ok	éliminé	éliminé	éliminé
2ème post-sélection		Ok	éliminé	Ok	Ok	Eliminé	Ok	Eliminé	éliminé	éliminé

		P21	P22	P23	P24	P25	P26	P27	P28	P29
Ext1	MPS 64k	3,10	-1,50	4,60	1,00	-6,00	-5,00	0,00	-7,00	-1,00
	MPS 96k	1,40	0,00	-0,90	0,00	-4,00	-4,00	-1,00	-5,00	-1,00
	MPS 128k	0,60	-1,50	-1,90	0,00	2,00	-1,00	0,00	-1,00	0,00
Ext2	MPS 64k	-0,40	-5,00	0,00	0,00	-5,00	-4,00	1,00	-3,00	-2,00
	MPS 96k	-0,40	-4,00	0,00	0,00	-5,00	-3,00	1,00	-4,00	-2,00
	MPS 128k	-0,40	-3,00	0,00	1,00	-3,00	-3,00	1,00	-1,00	-2,00
Ext3	MPS 64k	-0,60	0,00	0,00	0,00	-2,00	0,00	0,00	0,00	0,00
	MPS 96k	-1,50	-0,50	0,00	0,00	-5,00	1,00	0,00	0,00	1,00
	MPS 128k	0,00	-0,50	0,00	1,00	-2,00	0,00	0,00	0,00	1,00
Ext4	MPS 64k	-2,30	-4,00	0,20	0,00	-3,00	-3,00	-1,00	4,00	-1,00
	MPS 96k	-1,80	-3,00	-0,90	0,00	-3,00	-3,00	2,00	2,00	0,00
	MPS 128k	-1,60	0,00	0,20	1,00	-3,00	1,00	2,00	0,00	-1,00
Ext5	MPS 64k	0,00	-2,00	-0,10	1,00	-8,00	-4,00	-2,00	0,00	0,00
	MPS 96k	0,30	0,00	0,10	0,00	-3,00	-5,00	-4,00	2,50	-2,00
	MPS 128k	0,80	-1,00	0,10	0,00	-4,00	-2,00	-2,00	4,00	0,00
Ext6	MPS 64k	-0,30	1,50	-0,30	0,00	-3,00	-2,00	-1,00	-2,00	1,00
	MPS 96k	1,20	3,00	-0,40	1,00	-6,00	1,00	0,00	0,00	0,00
	MPS 128k	0,20	4,00	-0,20	0,00	-3,00	0,00	-1,00	-3,50	0,00
Ext7	MPS 64k	0,50	0,00	1,00	-2,00	-8,00	-6,00	-1,00	-3,00	-1,00
	MPS 96k	-0,30	-1,00	0,00	-2,00	-5,00	-2,00	-1,00	-2,00	0,00
	MPS 128k	-0,30	-1,00	1,00	-2,00	-5,00	-2,00	-1,00	2,00	-1,00
MOYENNE		-0,086	-0,929	0,119	0,000	-4,000	-2,190	-0,381	-0,810	-0,524
ECART-TYPE		1,198	2,193	1,194	0,949	2,191	2,112	1,396	2,853	0,981
P (m=0)		0,746	0,067	0,653	1	<0,0001	0	0,225	0,208	0,024
Post-sélection au seuil de 10%		éliminé	ok	éliminé	éliminé	ok	ok	éliminé	éliminé	ok
Post-sélection au seuil de 25%		éliminé	ok	éliminé	éliminé	ok	ok	ok	ok	ok
2ème post-sélection		éliminé	Ok	éliminé	éliminé	Ok	Ok	Ok	éliminé	Ok

ANNEXE C : 2ÈME TEST PERCEPTIF

RÉSULTATS DU 2ÈME TESTS

	Test 5.1											
	Ext 11 : opéra				Ext 12 : téléfilm				Ext 13 : documentaire			
Participant	Ext11 ref	Ext11 64k	Ext11 96k	Ext11 128k	Ext12 ref	Ext12 64k	Ext12 96k	Ext12 128k	Ext13 ref	Ext13 64k	Ext13 96k	Ext13 128k
1	10	8,8	10	10	10	10	7	10	10	8,8	10	10
2	10	5	4	9	7,6	10	10	6,5	10	5	6	7
3	10	6	4	9	6	10	8	8	8	8	8	10
4	10	7	9	9	10	10	10	10	9	7	10	10
5	10	7	8	9	10	10	10	10	10	6	9	9
6	10	3	2	4	5	10	4	7	10	1	6	8
7	8	10	8	10	10	7,5	7,2	-	7	6,8	-	10
8	10	5	6	7	10	5	8	6	10	6	8	8
9	7	6	10	9	9	9	9	10	10	9,5	7	8,5
MOYENNE	9,444	6,422	6,778	8,444	8,622	9,056	8,133	8,438	9,333	6,456	8,000	8,944
Écart-type (σ_{N-1})	1,130	2,099	2,906	1,878	1,958	1,740	1,952	1,761	1,118	2,497	1,604	1,130

	Test au casque (1/2)							
	Ext 21 : opéra				Ext 22 : téléfilm			
Participant	Ext21 d-mix	Ext21 64k	Ext21 96k	Ext21 128k	Ext22 d-mix	Ext22 64k	Ext22 96k	Ext22 128k
1	10	10	8	8	10	10	10	10
2	10	6,5	8	6	9	7	9	10
3	8	9	6	7	6	8	9	8
4	7	6	6	6	6	7	6	7
5	8	8	7	9	8	8	8,5	8
6	4	6	5	6	7	5	2	5
7	-	-	-	-	-	-	-	-
8	10	8	6	8	10	10	9	10
9	8	6	6	9	9	8,5	6	8
Moyenne	8,125	7,438	6,500	7,375	8,125	7,938	7,438	8,250
Écart-type (σ_{N-1})	2,031	1,545	1,069	1,302	1,642	1,657	2,638	1,753

	Test au casque (2/2)							
	Ext 23 : téléfilm				Ext 24 : documentaire			
Participant	Ext23 d-mix	Ext23 64k	Ext23 96k	Ext23 128k	Ext24 d-mix	Ext24 64k	Ext24 96k	Ext24 128k
1	9,9	10	10	9,8	10	10	10	10
2	10	9,6	8	7,5	9	10	10	7
3	9	10	7	10	7	7	7	7
4	7	8	7	7	7	7	7	7
5	9,5	10	10	9,5	8	8	8	8
6	7	6	3	3	7	3	3	7
7	7,1	8,2	7,5	10	8,2	6,5	10	9,4
8	8	10	10	9	10	9	9	9
9	10	8	7	9	7	8	9	7
Moyenne	8,611	8,867	7,722	8,311	8,133	7,611	8,111	7,933
Écart-type (σ_{N-1})	1,337	1,407	2,224	2,260	1,269	2,147	2,261	1,221

ANNEXE D : CONVERSION DES FLUX MPEG-4 ET MESURES DES PARAMÈTRES LOUDNESS

LIGNE DE COMMANDE UTILISÉE POUR CONVERTIR LES FLUX MPEG-4 DES PROGRAMMES ENTIERS EN FICHIERS STÉRÉOPHONIQUES EN PCM LINÉAIRE, À 48 KHZ EN 24 BITS

```
mbpro-207646:~ manuelnaudin$ ffmpeg -i /Volumes/MAC\
AN/R128/Prog_Enc_MPS-HEAAC/OperaIIA-P2_5.1_Enc_128kbps.m4a -acodec
pcm_s24le /Volumes/MAC\ AN/R128/Prog_Enc_MPS-HEAAC/OperaIIA-
P2_5.1_Enc_128kbps.wav
ffmpeg version 1.2 Copyright (c) 2000-2013 the FFmpeg developers
  built on Mar 31 2013 21:55:33 with Apple clang version 4.1 (tags/Apple/clang-
421.11.66) (based on LLVM 3.1svn)
  configuration: --prefix=/opt/local --enable-swscale --enable-avfilter --enable-libmp3lame
--enable-libvorbis --enable-libopus --enable-libtheora --enable-libschroedinger --enable-
libopenjpeg --enable-libmodplug --enable-libvpx --enable-libspeex --enable-libfreetype --
mandir=/opt/local/share/man --enable-shared --enable-pthreads --cc=/usr/bin/clang --
arch=x86_64 --enable-yasm --enable-gpl --enable-postproc --enable-libx264 --enable-
libxvid
  libavutil      52. 18.100 / 52. 18.100
  libavcodec     54. 92.100 / 54. 92.100
  libavformat    54. 63.104 / 54. 63.104
  libavdevice    54.  3.103 / 54.  3.103
  libavfilter     3. 42.103 /  3. 42.103
  libswscale      2.  2.100 /  2.  2.100
  libswresample   0. 17.102 /  0. 17.102
  libpostproc    52.  2.100 / 52.  2.100
Input #0, mov,mp4,m4a,3gp,3g2,mj2, from '/Volumes/MAC
AN/R128/Prog_Enc_MPS-HEAAC/OperaIIA-P2_5.1_Enc_128kbps.m4a':
  Metadata:
    major_brand     : mp42
    minor_version   : 0
    compatible_brands: mp42isom
    creation_time   : 2013-04-11 12:34:35
  Duration: 01:15:44.48, start: 0.190688, bitrate: 129 kb/s
```

```
    Stream #0:0(und): Audio: aac (mp4a / 0x6134706D), 48000 Hz, stereo, fltp, 127 kb/s
    Metadata:
      creation_time   : 2013-04-11 12:34:35
      handler_name    : soun
Output #0, wav, to '/Volumes/MAC AN/R128/Prog_Enc_MPS-HEAAC/OperaIIA-
P2_5.1_Enc_128kbps.wav':
  Metadata:
    major_brand     : mp42
    minor_version   : 0
    compatible_brands: mp42isom
    ISFT            : Lavf54.63.104
    Stream #0:0(und): Audio: pcm_s24le ([1][0][0][0] / 0x0001), 48000 Hz, stereo, s32,
2304 kb/s
    Metadata:
      creation_time   : 2013-04-11 12:34:35
      handler_name    : soun
Stream mapping:
  Stream #0:0 -> #0:0 (aac -> pcm_s24le)
Press [q] to stop, [?] for help
size= 1278192kB time=01:15:44.68 bitrate=2304.0kbits/s
video:0kB audio:1278192kB subtitle:0 global headers:0kB muxing overhead 0.000008%
mbpro-207646:~ manuelnaudin$
```