

Union internationale des télécommunications

UIT-R

Secteur des Radiocommunications de l'UIT

Recommandation UIT-R BS.1196-2
(03/2010)

**Codage audio pour la
radiodiffusion numérique**

Série BS
Service de radiodiffusion sonore



Union
internationale des
télécommunications

Avant-propos

Le rôle du Secteur des radiocommunications est d'assurer l'utilisation rationnelle, équitable, efficace et économique du spectre radioélectrique par tous les services de radiocommunication, y compris les services par satellite, et de procéder à des études pour toutes les gammes de fréquences, à partir desquelles les Recommandations seront élaborées et adoptées.

Les fonctions réglementaires et politiques du Secteur des radiocommunications sont remplies par les Conférences mondiales et régionales des radiocommunications et par les Assemblées des radiocommunications assistées par les Commissions d'études.

Politique en matière de droits de propriété intellectuelle (IPR)

La politique de l'UIT-R en matière de droits de propriété intellectuelle est décrite dans la «Politique commune de l'UIT-T, l'UIT-R, l'ISO et la CEI en matière de brevets», dont il est question dans l'Annexe 1 de la Résolution UIT-R 1. Les formulaires que les titulaires de brevets doivent utiliser pour soumettre les déclarations de brevet et d'octroi de licence sont accessibles à l'adresse <http://www.itu.int/ITU-R/go/patents/fr>, où l'on trouvera également les Lignes directrices pour la mise en oeuvre de la politique commune en matière de brevets de l'UIT-T, l'UIT-R, l'ISO et la CEI et la base de données en matière de brevets de l'UIT-R.

Séries des Recommandations UIT-R

(Egalement disponible en ligne: <http://www.itu.int/publ/R-REC/fr>)

Séries	Titre
BO	Diffusion par satellite
BR	Enregistrement pour la production, l'archivage et la diffusion; films pour la télévision
BS	Service de radiodiffusion sonore
BT	Service de radiodiffusion télévisuelle
F	Service fixe
M	Services mobile, de radiorepérage et d'amateur y compris les services par satellite associés
P	Propagation des ondes radioélectriques
RA	Radio astronomie
RS	Systèmes de télédétection
S	Service fixe par satellite
SA	Applications spatiales et météorologie
SF	Partage des fréquences et coordination entre les systèmes du service fixe par satellite et du service fixe
SM	Gestion du spectre
SNG	Reportage d'actualités par satellite
TF	Emissions de fréquences étalon et de signaux horaires
V	Vocabulaire et sujets associés

Note: Cette Recommandation UIT-R a été approuvée en anglais aux termes de la procédure détaillée dans la Résolution UIT-R 1.

Publication électronique
Genève, 2011

© UIT 2011

Tous droits réservés. Aucune partie de cette publication ne peut être reproduite, par quelque procédé que ce soit, sans l'accord écrit préalable de l'UIT.

RECOMMANDATION UIT-R BS.1196-2*, **

Codage audio pour la radiodiffusion numérique

(Question UIT-R 19/6)

(1995-2001-2010)

Domaine d'application

La présente Recommandation spécifie les systèmes de codage des sources audio applicables à la radiodiffusion sonore et visuelle numérique. Elle fournit également les spécifications d'un système destiné à l'amélioration multicanal rétrocompatible de systèmes de radiodiffusion sonore et visuelle numérique.

L'Assemblée des radiocommunications de l'UIT,

considérant

- a) que les spécifications utilisateur en matière de systèmes de codage audio pour la radiodiffusion numérique sont énoncées dans la Recommandation UIT-R BS.1548;
- b) que la Recommandation UIT-R BS.775 décrit un système sonore multicanal avec ou sans image associée et qu'un système sonore multicanal de haute qualité utilisant une réduction efficace du débit binaire est un élément essentiel des systèmes de radiodiffusion numérique;
- c) que l'évaluation subjective des dégradations faibles dans les systèmes audio, y compris les systèmes audio multivoies, fait l'objet de la Recommandation UIT-R BS.1116;
- d) que l'évaluation subjective du niveau de qualité sonore intermédiaire des systèmes audio fait l'objet de la Recommandation UIT-R BS.1534 (MUSHRA);
- e) que le Secteur des radiocommunications de l'UIT a testé un système de codage à faible débit binaire pour des signaux audio de haute qualité;
- f) que l'usage de techniques de codage de source audio similaires pour différents services pourrait permettre une plus grande souplesse d'utilisation des systèmes et une diminution du coût des récepteurs;
- g) que plusieurs services de radiodiffusion utilisent déjà des codecs audio appartenant aux familles MPEG-1, MPEG-2, MPEG-4, AC-3 et E-AC-3, ou en ont spécifié l'usage;
- h) que la Recommandation UIT-R R BS.1548 fournit une liste de codecs dont on a démontré la conformité aux spécifications du radiodiffuseur pour la contribution, la distribution et l'émission;
- j) que les radiodiffuseurs qui n'ont pas encore démarré de services doivent être à même de choisir le système le mieux adapté à leur application;
- k) que les radiodiffuseurs peuvent avoir besoin d'étudier la compatibilité avec les systèmes et équipements de radiodiffusion hérités lorsqu'ils choisissent un nouveau système;
- l) que lors de la mise en place d'un système audio multicanal, il convient de prendre en compte les récepteurs monophoniques et stéréophoniques existants;

* La Commission d'études 6 des radiocommunications a apporté des modifications rédactionnelles à cette Recommandation en 2003 conformément à la Résolution UIT-R 44.

** Cette Recommandation doit être portée à l'attention de l'Organisation internationale de standardisation (ISO) et de la Commission électrotechnique internationale (CEI).

m) qu'une extension multicanal rétrocompatible réalisée sur un système de codage audio existant peut apporter une meilleure efficacité en termes de débit binaire que la radiodiffusion simultanée,

recommande

1 d'utiliser, pour les nouvelles applications de radiodiffusion sonore ou télévisuelle numérique, lorsque la compatibilité avec les transmissions et les équipements hérités n'est pas requise, l'un des systèmes de codage audio à faible débit binaire suivants:

- MPEG-4 HE-AAC v2, tel que spécifié dans le Document ISO/CEI 14496-3:2009;
- E-AC-3, tel que spécifié dans le Document ETSI TS 102 366 (2008-08).

NOTE 1 – MPEG-4 HE-AAC v2 et E-AC-3 sont des surensembles plus souples des systèmes MPEG-4 AAC-LC et AC-3.

2 d'utiliser, pour les applications de radiodiffusion sonore ou télévisuelle numérique, lorsque la compatibilité avec les transmissions et les équipements hérités est requise, l'un des systèmes de codage à faible débit binaire suivants:

- MPEG-1 de couche II, tel que spécifié dans le Document ISO/CEI 11172-3:1993;
- MPEG-2 de couche II à fréquence d'échantillonnage moitié, tel que spécifié dans le Document ISO/CEI 13818-3:1998;
- MPEG-2 AAC-LC ou MPEG-2 AAC-LC avec SBR, tel que spécifié dans le Document ISO/CEI 13818-7:2006;
- MPEG-4 AAC-LC, tel que spécifié dans le Document ISO/CEI 14496-3:2009;
- MPEG-4 HE AAC v2, tel que spécifié dans le Document ISO/CEI 14496-3:2009;
- AC-3, tel que spécifié dans le Document ETSI TS 102 366 (2008-08).

NOTE 1 – Le codec ISO/CEI 11172-3 peut parfois être désigné sous le nom de codec 13818-3 étant donné que la spécification de ce dernier se réfère à la Norme 11172-3.

3 d'utiliser, pour l'extension multicanal rétrocompatible des systèmes de radiodiffusion sonore et télévisuelle numérique, les extensions audio multicanal décrites dans le Document ISO/CEI 23003-1:2007;

NOTE 1 – Etant donné que la technologie MPEG ambiophonique décrite dans le Document ISO/CEI 23003-1:2007 est indépendante de la technologie de compression (codeur de base) utilisée pour la transmission du signal rétrocompatible, les outils d'amélioration multicanal décrits peuvent être utilisés en combinaison avec l'un quelconque des systèmes de codage recommandés aux *recommande* 1 et 2.

4 d'utiliser éventuellement, pour les liaisons de distribution et de contribution, le système de codage de couche II de l'ISO/CEI 11172-3, avec un débit binaire d'au moins 180 kbit/s par signal audio (c'est-à-dire par signal mono ou par composante d'un signal stéréo dont les deux voies ont été codées indépendamment), à l'exclusion des données auxiliaires;

5 d'utiliser éventuellement, pour les liaisons de diffusion de commentaires, le système de codage de couche III de l'ISO/CEI 11172-3, avec un débit binaire d'au moins 60 kbit/s, à l'exclusion des données auxiliaires, pour les signaux mono, et un débit d'au moins 120 kbit/s, à l'exclusion des données auxiliaires, pour les signaux stéréo, en utilisant un codage stéréo combiné;

6 d'utiliser une fréquence d'échantillonnage de 48 kHz pour les applications qui requièrent une haute qualité;

7 de ne pas appliquer d'accentuation au signal d'entrée du codeur audio à faible débit binaire, le codeur ne devant pas, lui non plus, en introduire;

8 que le respect de la présente Recommandation se fasse à titre volontaire. Cela étant, la Recommandation peut contenir certaines dispositions obligatoires (pour assurer, par exemple, l'interopérabilité et l'applicabilité) et elle est considérée comme respectée lorsque toutes ces dispositions sont observées. Le futur d'obligation et les autres moyens d'expression de l'obligation comme le verbe «devoir» ainsi que leurs formes négatives servent à énoncer des prescriptions. L'utilisation de ces formes ne signifie aucunement qu'il est obligatoire de respecter en partie ou en totalité la présente Recommandation,

recommande en outre

1 de faire référence à la Recommandation UIT-R BS.1548 pour toute information concernant les configurations des systèmes de codage qui se sont révélés être conformes aux prescriptions de qualité et autres pour la contribution, la distribution et l'émission.

NOTE 1 – Les Appendices 1 à 4 fournissent des informations sur les codecs mentionnés dans la présente Recommandation.

Appendice 1

MPEG-1 et MPEG-2, couches II et III audio

1 Codage

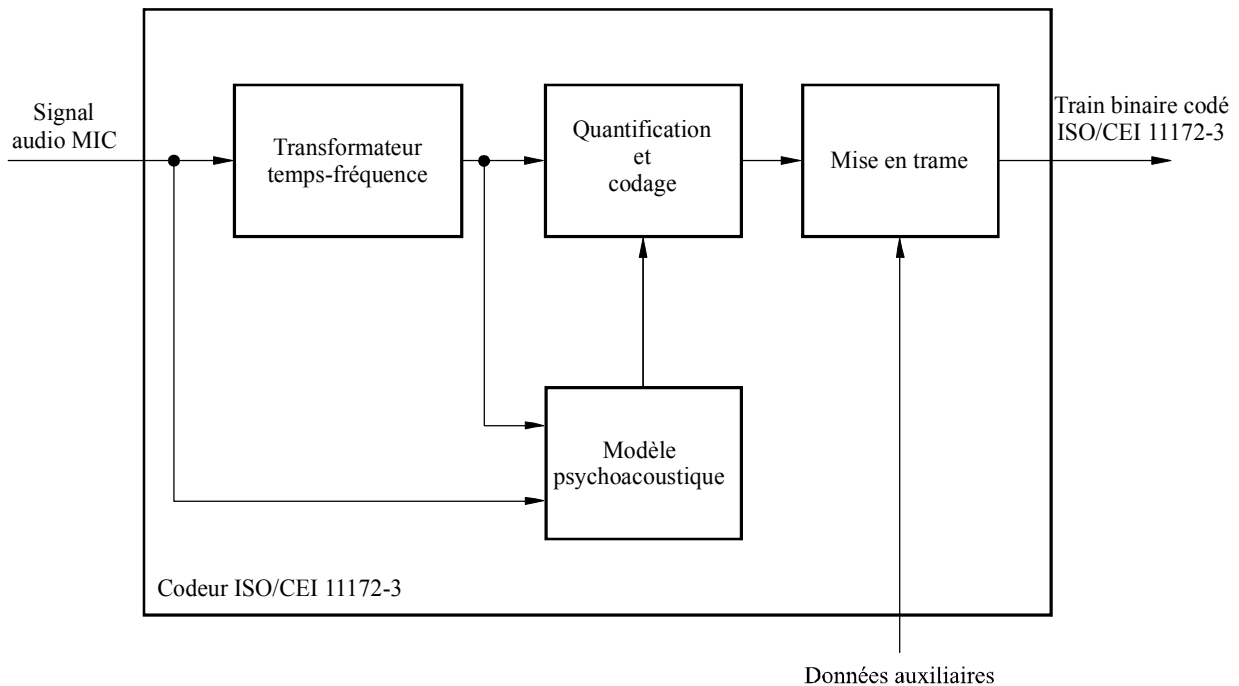
Le codeur traite le signal audionumérique et comprime le train binaire. L'algorithme de codage n'est pas normalisé et peut utiliser divers moyens pour effectuer le codage, par exemple l'estimation du seuil de masquage à l'écoute, la quantification et les facteurs d'échelle (Note 1 ci-après). Toutefois, la sortie du codeur doit être telle qu'un décodeur conforme à la présente Recommandation puisse produire un signal audio convenant pour l'application souhaitée.

NOTE 1 – On obtiendra une norme minimale de qualité satisfaisante avec un codeur conforme à la description donnée dans les Annexes C et D de la Norme ISO/CEI 11172-3, 1993.

La Fig. 1 décrit un codeur type. Le codeur est alimenté par des échantillons de signaux audio. La transformation temps-fréquence donne une représentation filtrée et sous-échantillonnée du flux audio entrant. Les échantillons transformés peuvent être appelés échantillons de sous-bande (comme dans la couche I ou II, voir ci-après) ou échantillons de sous-bande transformés (comme dans la couche III). Un modèle psychoacoustique, utilisant une transformée de Fourier rapide, parallèlement à la transformation temps-fréquence du signal audio, crée un ensemble de données permettant de commander la quantification et le codage. Ces données sont différentes selon les modalités concrètes de mise en œuvre du codeur. Une possibilité consiste à utiliser une évaluation du seuil de masquage pour commander le quantificateur. Le bloc de quantification et de codage génère un ensemble de symboles de codage à partir des échantillons entrants transformés. Là aussi, la fonction de transfert de ce bloc dépend du système de codage. Le bloc «mise en trame» assemble, pour la couche choisie, le train binaire proprement dit à partir des données de sortie des autres blocs (données d'allocation des bits, facteurs d'échelle, échantillons de sous-bande codés, etc.) et ajoute d'autres informations dans le champ des données auxiliaires (protection contre les erreurs par exemple), si nécessaire.

FIGURE 1

Schéma fonctionnel d'un codeur type



BS.1196-01

2 Couches

Selon l'application, on peut utiliser différentes couches du système de codage: la complexité du codeur s'accroît et sa qualité s'améliore en fonction de la couche utilisée.

Couche I: Cette couche comprend la transformation de base des signaux audionumériques entrants en 32 sous-bandes, une segmentation fixe pour formater les données en blocs, un modèle psychoacoustique permettant de déterminer l'allocation adaptative des bits et une quantification utilisant la compression-extension et le formatage des blocs. Une trame de la couche I représente 384 échantillons par voie.

Couche II: Cette couche apporte un codage de l'allocation des bits, des facteurs d'échelle et des échantillons supplémentaires. Une trame de la couche II représente $3 \times 384 = 1\,152$ échantillons par voie.

Couche III: Cette couche apporte une résolution de fréquence plus élevée, qui repose sur un banc de filtres hybride (banc de filtres à 32 sous-bandes avec transformée en cosinus discrète modifiée de longueur variable). Elle ajoute une quantification non uniforme, une segmentation adaptative et un codage entropique des valeurs quantifiées. Une trame de la couche III représente 1 152 échantillons par voie.

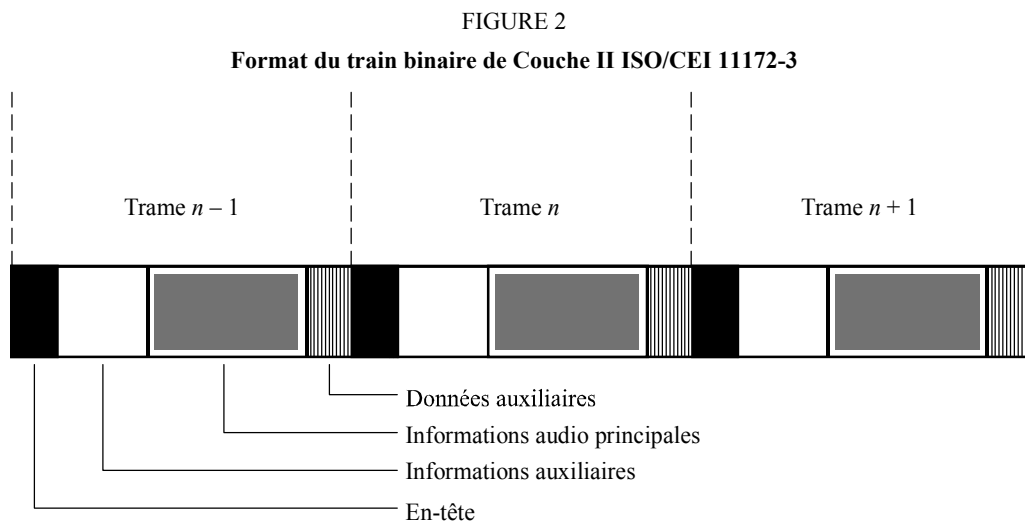
Quatre modes différents sont possibles pour chacune des couches:

- voie unique;
- voie double (deux signaux audio indépendants codés à l'intérieur d'un seul train binaire, application bilingue par exemple);
- stéréo (les signaux gauche et droit d'une paire stéréophonique sont codés à l'intérieur d'un même train binaire);

- stéréo combiné (les signaux gauche et droit d'une paire stéréophonique sont codés dans un seul et même train binaire en exploitant la redondance et la présence d'éléments non pertinents pour la stéréo). Le mode stéréo combiné peut servir à améliorer la qualité audio aux faibles débits binaires et/ou à réduire le débit binaire des signaux stéréophoniques.

3 Format du train binaire codé

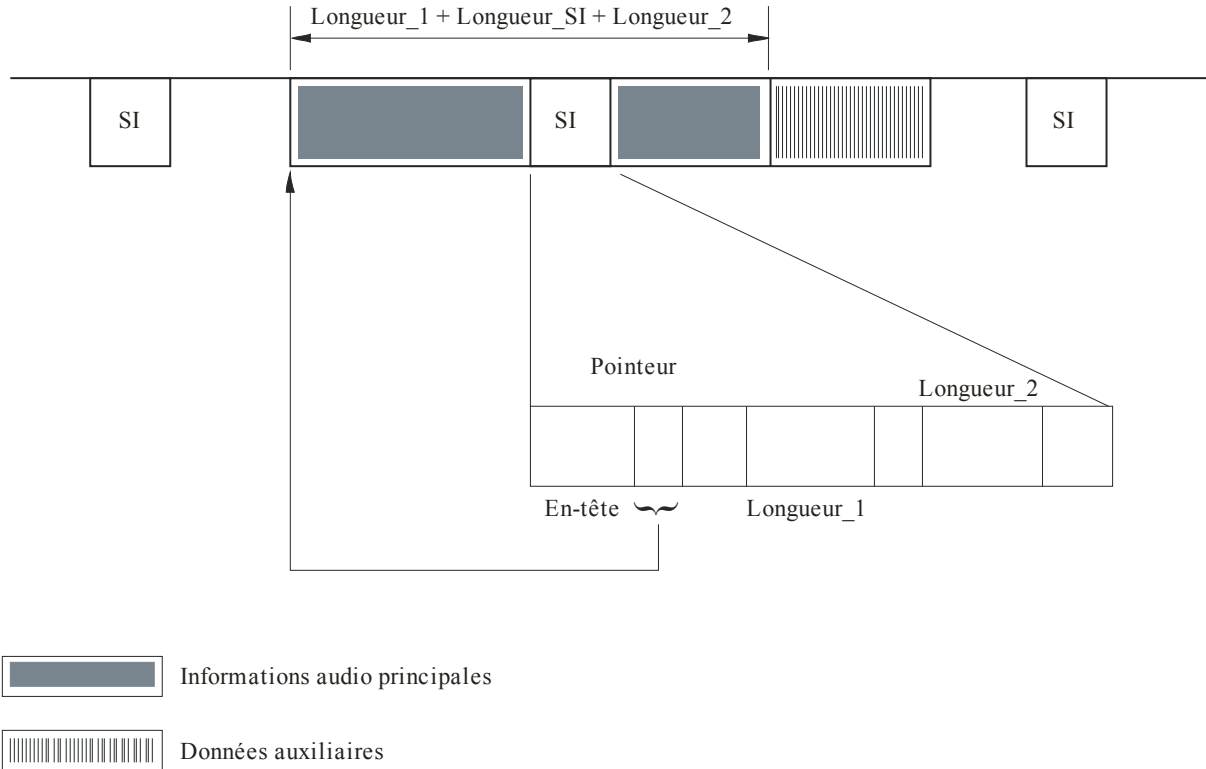
Les Fig. 2 et 3 donnent, respectivement pour la couche II et la couche III, une description schématique du train binaire ISO/CEI 11172-3. Un train binaire codé est composé de trames successives. Les champs contenus dans la trame diffèrent selon la couche, comme indiqué ci-dessous:



Couche II:

En-tête:	partie du train binaire contenant les informations de synchronisation et d'état
Informations auxiliaires:	partie du train binaire contenant les informations sur l'allocation des bits et les facteurs d'échelle
Informations audio principales:	partie du train binaire contenant les échantillons de sous-bande codés
Données auxiliaires:	partie du train binaire contenant des données définissables par l'utilisateur

FIGURE 3
Format du train binaire de Couche III ISO/CEI 11172-3



Couche III:

Informations auxiliaires (SI):	partie du train binaire contenant l'en-tête, le pointeur, la longueur_1 et la longueur_2, les informations sur les facteurs d'échelle, etc.
En-tête:	partie du train binaire contenant les informations de synchronisation et d'état
Pointeur:	indique le début des informations audio principales
Longueur_1:	longueur de la première partie des informations audio principales
Longueur_2:	longueur de la seconde partie des informations sur les principales données audio
Information audio principales:	partie du train binaire contenant les signaux audio codés
Données auxiliaires:	partie du train binaire contenant des données définissables par l'utilisateur

BS.1196-03

4 Décodage

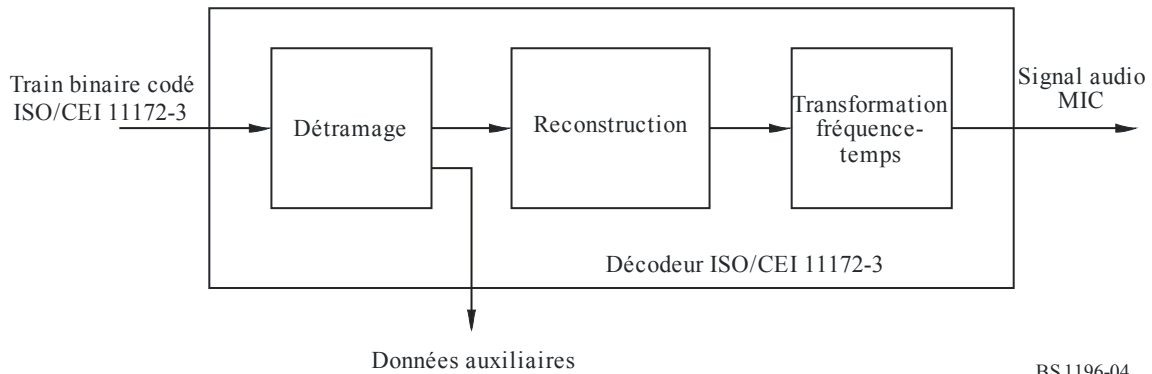
Le décodeur accepte les trains binaires audio comprimés dans la syntaxe définie dans la Norme ISO/CEI 11172-3, décode les éléments de données et utilise les informations pour produire le signal audionumérique de sortie.

Le décodeur est alimenté par le train binaire audio codé. Le mécanisme de déramage et de décodage du train binaire effectue en option une détection d'erreur si le codeur applique un mécanisme de correction d'erreur. Le train binaire est déramé de manière à retrouver les différents éléments d'information comme l'en-tête de trame audio, l'allocation des bits, les facteurs d'échelle, les échantillons transformés et, éventuellement, les données auxiliaires. Le processus de

reconstruction reconstruit la version quantifiée de l'ensemble des échantillons transformés. La transformation fréquence-temps retransforme ces échantillons en signaux audio au format MIC linéaire.

FIGURE 4

Schéma fonctionnel du décodeur



BS.1196-04

Appendice 2

MPEG-2 et MPEG-4 AAC audio

1 Introduction

La Norme ISO/CEI 13818-7 décrit la norme MPEG-2 audio non rétrocompatible, désignée sous le nom de codage audio avancé (AAC) MPEG-2, norme applicable à la diffusion multicanal de qualité supérieure à celle que l'on obtient lorsque la rétrocompatibilité avec la Norme MPEG-1 est imposée.

Le système AAC définit trois profils permettant d'obtenir un compromis entre, d'une part, la mémoire et la puissance de traitement nécessaires et, d'autre part, la qualité audio, à savoir:

– *Profil principal*

Le profil principal offre la qualité audio la plus élevée quel que soit le débit. Tous les outils à l'exception de la commande de gain peuvent être utilisés pour produire une qualité audio élevée. La mémoire et la puissance de traitement nécessaires sont supérieures à celles requises par le profil faible complexité (LC, *low complexity*). Un décodeur utilisant le profil principal peut décoder un train binaire codé selon le profil LC.

– *Profil faible complexité (LC, low complexity)*

La puissance de traitement et la mémoire nécessaires au profil LC sont inférieures à celles requises pour le profil principal, mais la qualité reste élevée. Dans le profil LC, ni la prédiction ni l'outil de commande de gain ne sont autorisés et la commande de mise en forme du bruit dans le domaine temporel (TNS, temporal noise shaping) est limitée.

– *Profil taux d'échantillonnage adaptable (SSR, scalable sampling rate)*

Le profil SSR permet d'offrir un signal adaptable en fréquence avec outil de commande de gain. Il permet de choisir certaines bandes de fréquences à décoder, ce qui permet de limiter les éléments matériels du décodeur. Par exemple, pour décoder seulement la bande de fréquences la plus basse à la fréquence d'échantillonnage de 48 kHz, le décodeur peut reproduire le signal audio d'une largeur de bande de 6 kHz avec une complexité de décodage minimale.

Le système AAC prend en charge 12 types de fréquences d'échantillonnage comprises entre 8 et 96 kHz (voir le Tableau 1), et un maximum de 48 canaux audio. Le Tableau 2 présente les configurations de canaux par défaut, à savoir les configurations monocanal, bicanal, à cinq canaux (trois à l'avant/deux à l'arrière) et à cinq canaux plus un canal à effets basses fréquences (LFE, *low-frequency effects*) (largeur de bande < 200 Hz), etc. Outre ces configurations par défaut, il est possible de spécifier le nombre de haut-parleurs à chaque emplacement (à l'avant, sur les côtés et à l'arrière), ce qui autorise une certaine souplesse dans la disposition multicanal des haut-parleurs. La réduction par mixage du nombre de canaux («*down-mix*») est également prise en charge. L'utilisateur peut désigner un coefficient de réduction pour effectuer une conversion multicanal/bicanal. Il est par conséquent possible de régler la qualité du son au moyen d'un dispositif de lecture comportant seulement deux canaux.

TABLEAU 1

Fréquences d'échantillonnage prises en charge

Fréquence d'échantillonnage (Hz)
96 000
88 200
64 000
48 000
44 100
32 000
24 000
22 050
16 000
12 000
11 025
8 000

TABLEAU 2

Configurations de canaux par défaut

Nombre de haut-parleurs	Éléments de syntaxe audio, énumérés par ordre de réception	Correspondance par défaut entre l'élément de syntaxe et le(s) haut-parleur(s)
1	single_channel_element	Haut-parleur avant central
2	channel_pair_element	Haut-parleurs avant gauche et droit

TABLEAU 2 (*fin*)

Nombre de haut-parleurs	Éléments de syntaxe audio, énumérés par ordre de réception	Correspondance par défaut entre l'élément de syntaxe et le(s) haut-parleur(s)
3	single_channel_element()	Haut-parleur avant central
	channel_pair_element()	Haut-parleurs avant gauche et droit
4	single_channel_element()	Haut-parleur avant central
	channel_pair_element(),	Haut-parleurs avant gauche et droit
	single_channel_element()	Haut-parleur ambiophonique arrière
5	single_channel_element()	Haut-parleur avant central
	channel_pair_element()	Haut-parleurs avant gauche et droit
	channel_pair_element()	Haut-parleurs arrière ambiophoniques gauche et droit
5 + 1	single_channel_element()	Haut-parleur avant central
	channel_pair_element()	Haut-parleurs avant gauche et droit
	channel_pair_element()	Haut-parleurs arrière ambiophoniques gauche et droit
	lfe_element()	Caisson de basses
7 + 1	single_channel_element()	Haut-parleur avant central
	channel_pair_element()	Haut-parleurs avant gauche et droit
	channel_pair_element()	Haut-parleurs avant externes gauche et droit
	channel_pair_element()	Haut-parleurs arrière ambiophoniques gauche et droit
	lfe_element()	Caisson de basses

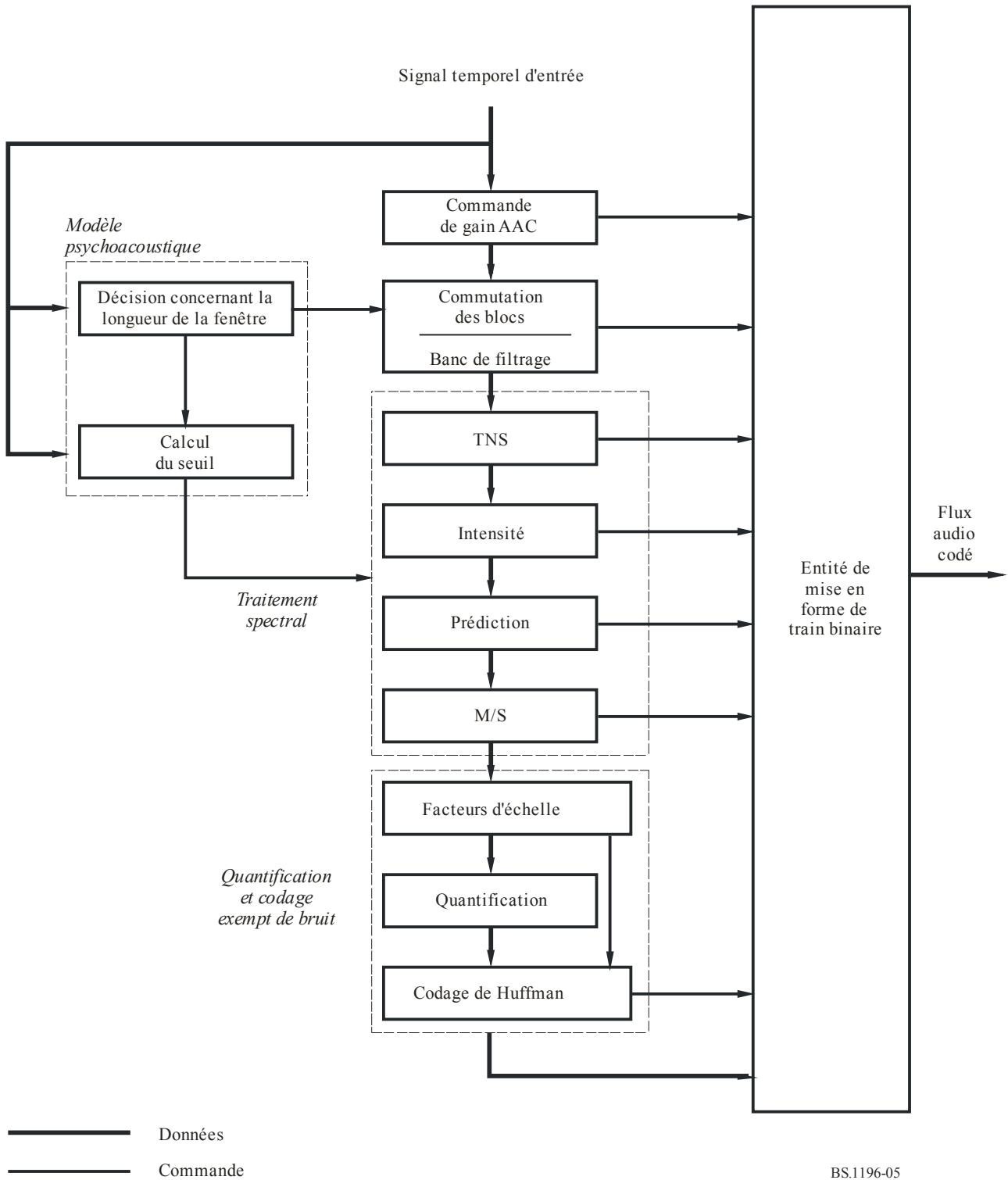
2 Codage

La structure de base du codeur MPEG-2 AAC est représentée à la Fig. 5. Le système AAC est composé des outils de codage suivants:

- *Commande de gain*: Commande partageant le signal d'entrée en quatre bandes de fréquences espacées de façon égale. La commande de gain est utilisée pour le profil SSR.
- *Banc de filtres*: Dispositif (TCDM, transformée en cosinus discrète modifiée) décomposant le signal d'entrée en composantes spectrales sous-échantillonnées avec une résolution en fréquence de 23 Hz et une résolution temporelle de 21,3 ms (128 composantes spectrales) ou avec une résolution en fréquence de 187 Hz et une résolution temporelle de 2,6 ms (1 024 composantes spectrales) à une fréquence d'échantillonnage de 48 kHz. On choisit la taille de la fenêtre parmi deux tailles de fenêtre possibles.
- *Mise en forme du bruit dans le domaine temporel (TNS, temporal noise shaping)*: L'opération TNS suit l'étape du banc de filtres d'analyse. La technique TNS permet au codeur de contrôler la structure fine temporelle du bruit de quantification.

- *Codage stéréo M/S (mid/side) et codage stéréo intensité:* Dans le cas de signaux audio multicanaux, le codage stéréo intensité et le codage stéréo M/S peuvent être appliqués. Dans le codage stéréo intensité, seule l'enveloppe spectrale est transmise de manière à réduire les informations de direction qui sont émises. Dans le codage stéréo M/S, on peut transmettre la somme normalisée (M pour *middle*) et les signaux de différence (S pour *side*) au lieu des signaux gauche et droit d'origine.
- *Prédiction:* Pour réduire la redondance des signaux stationnaires, on effectue la prédiction du domaine temporel entre les composantes spectrales sous-échantillonnées des trames successives.
- *Quantification et codage exempt de bruit:* L'outil de quantification repose sur l'utilisation d'un quantificateur non uniforme avec un pas de 1,5 dB. Le codage de Huffman s'applique au spectre quantifié, aux différents facteurs d'échelle et aux informations de direction.
- *Entité de mise en forme de train binaire:* On utilise enfin une entité de mise en forme de train binaire pour multiplexer le train binaire, qui se compose des coefficients spectraux quantifiés et codés et d'informations supplémentaires provenant de chaque outil.
- *Modèle psychoacoustique:* On calcule le seuil de masquage courant au moyen d'un modèle psychoacoustique à partir du signal d'entrée. Le modèle psychoacoustique utilisé est analogue au modèle psychoacoustique 2 de la Norme ISO/CEI 11172-3. On emploie un rapport signal/masquage, calculé à partir du seuil de masquage et du niveau du signal d'entrée, au cours du processus de quantification afin de limiter le bruit de quantification audible et de choisir un outil de codage approprié.

FIGURE 5
Schéma fonctionnel d'un codeur MPEG-2 AAC

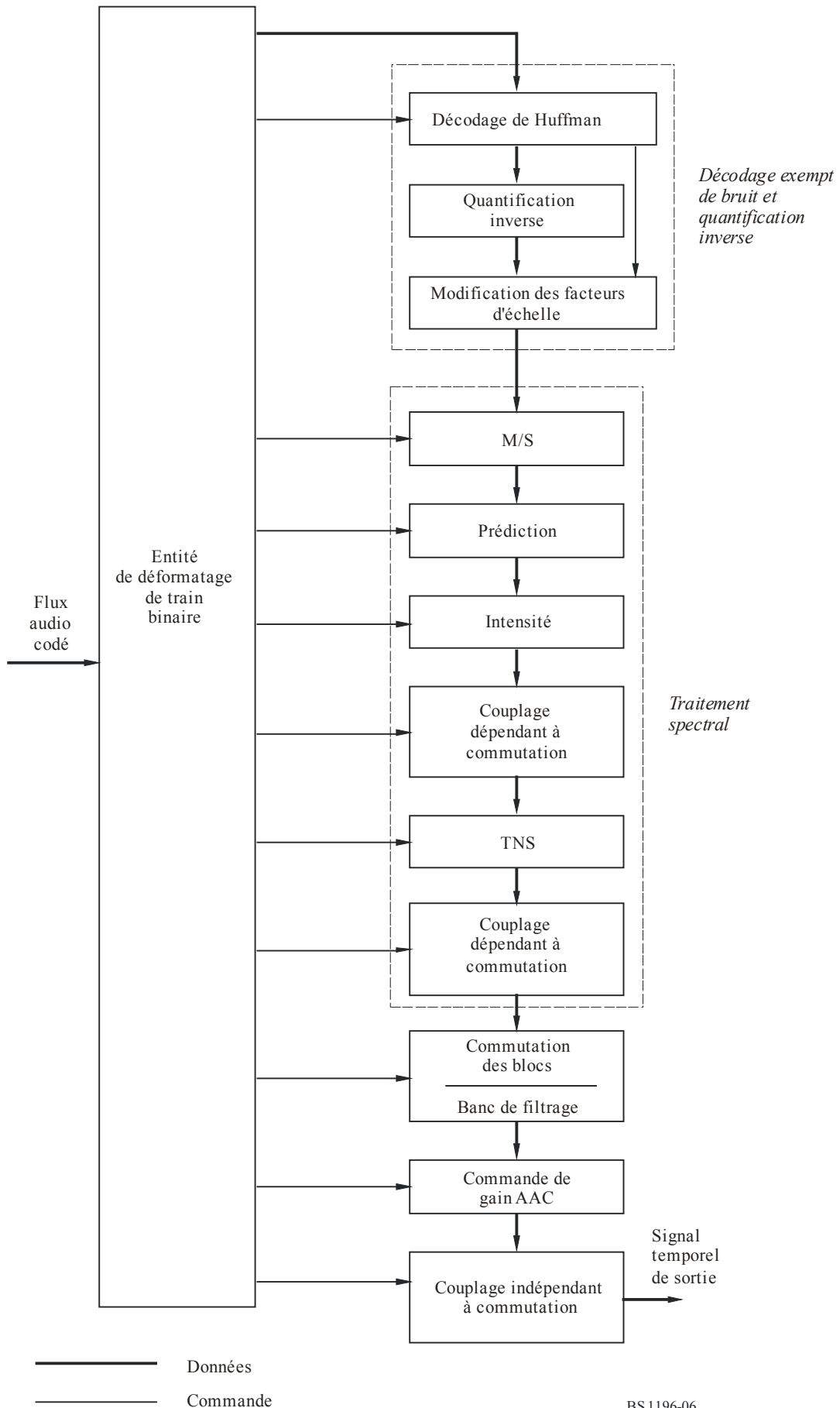


3 Décodage

La structure de base du décodeur MPEG-2 AAC est représentée à la Fig. 6. Le processus de décodage consiste essentiellement en l'inverse du processus de codage.

FIGURE 6

Schéma fonctionnel d'un décodeur MPEG-2 AAC

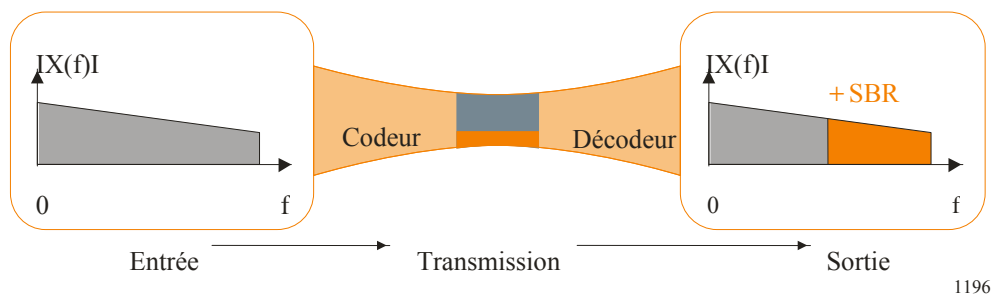


Le décodeur a pour fonctions de trouver la description des spectres audio quantifiés dans le train binaire, de décoder les valeurs quantifiées et les autres informations relatives à la reconstruction, de reconstruire les spectres quantifiés, de traiter les spectres reconstruits au moyen des outils actifs dans le train binaire de manière à obtenir les spectres du signal réel tels que décrits par le train binaire d'entrée et, enfin, de convertir le domaine spectral en domaine temporel, avec ou sans outil optionnel de commande de gain. Après reconstruction initiale et mise à l'échelle du spectre reconstruit, on peut utiliser de nombreux outils optionnels pour modifier un ou plusieurs spectres afin de renforcer l'efficacité du codage. Pour chacun des outils optionnels fonctionnant dans le domaine spectral, l'option «traverser» (*pass through*) est retenue et, dans tous les cas où une opération spectrale est omise, les spectres en entrée traversent directement l'outil sans modification.

4 HE AAC et reconstruction de bande spectrale

Le profil HE AAC (*High Efficiency AAC*, AAC à haute efficacité) apporte la reconstruction de bande spectrale (SBR, *spectral band replication*). La SBR est une méthode très efficace de codage des hautes fréquences utilisée dans les algorithmes de compression audio. Elle renforce l'efficacité des codecs audio et vocaux de faible débit binaire en augmentant la largeur de bande audio à un débit donné ou en améliorant l'efficacité du codage à un niveau de qualité donné.

Seule est codée et émise la partie inférieure du spectre, celle à laquelle l'oreille humaine est la plus sensible. Au lieu de transmettre la partie supérieure du spectre, ce système utilise la SBR comme un processus de post-décodage pour reconstruire les fréquences supérieures à partir d'une analyse des fréquences inférieures transmises. Pour garantir l'exactitude de la reconstruction, des paramètres utilisés par la SBR sont transmis dans le train binaire à un débit très faible.



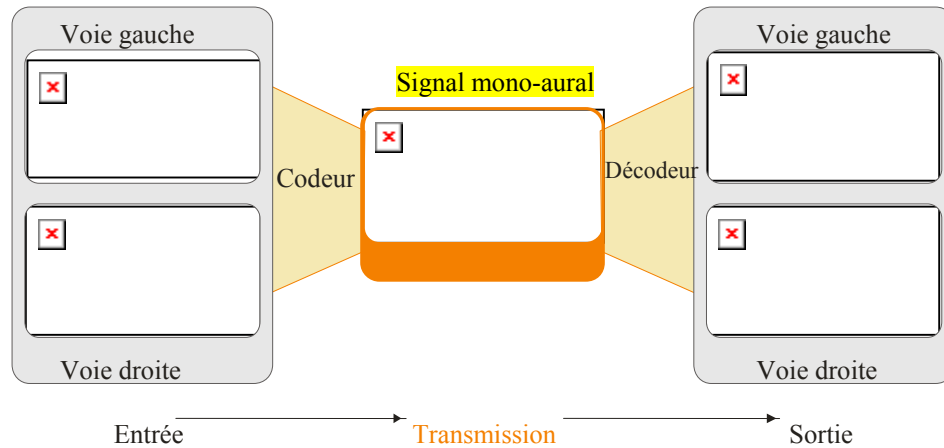
1196

Le train binaire HE AAC est une version améliorée du train binaire audio AAC. Les données SBR supplémentaires sont insérées dans l'élément de remplissage AAC (élément «FIL»), ce qui garantit une compatibilité avec la norme AAC. La technologie HE AAC est un système à débit dual. Le train binaire audio AAC pur rétrocompatible est transmis à une fréquence d'échantillonnage égale à la moitié de celle correspondant aux améliorations SBR. Ainsi, un décodeur AAC, qui n'est pas conçu pour décoder les données d'amélioration SBR, produit en sortie un signal temporel dont la fréquence d'échantillonnage est égale à la moitié de celle produite par un décodeur HE AAC.

5 HE AAC version 2 et stéréo paramétrique

Le HE AAC v2 est une extension du HE AAC qui offre la stéréo paramétrique (PS, *parametric stereo*) en vue de renforcer l'efficacité de la compression audio pour les signaux stéréo à faible débit binaire.

Le codeur analyse le signal audio stéréo et construit une représentation paramétrique de l'image stéréo. Il n'est dès lors plus nécessaire de transmettre les deux canaux et seule une représentation mono-aurale du signal stéréo d'origine est codée. Les paramètres nécessaires à la reconstruction de l'image stéréo sont transmis avec le signal.



1196

En conséquence, la qualité audio perçue d'un train binaire de faible débit (par exemple, 24 kbit/s) intégrant une stéréo paramétrique est bien supérieure à la qualité d'un train binaire analogue sans stéréo paramétrique.

Le train binaire HE AAC v2 est construit sur le train binaire HE AAC. Les données supplémentaires pour la stéréo paramétrique sont insérées dans l'élément d'extension SBR du train HE AAC mono, ce qui garantit une compatibilité avec le HE AAC et l'AAC.

Un décodeur HE AAC, qui n'est pas conçu pour décoder les données d'amélioration liées à la stéréo paramétrique, produit un signal de sortie monophonique à pleine largeur de bande. Un décodeur AAC pur, qui n'est pas conçu pour décoder les données d'amélioration SBR, produit en sortie un signal temporel à la moitié de la fréquence d'échantillonnage.

Appendice 3

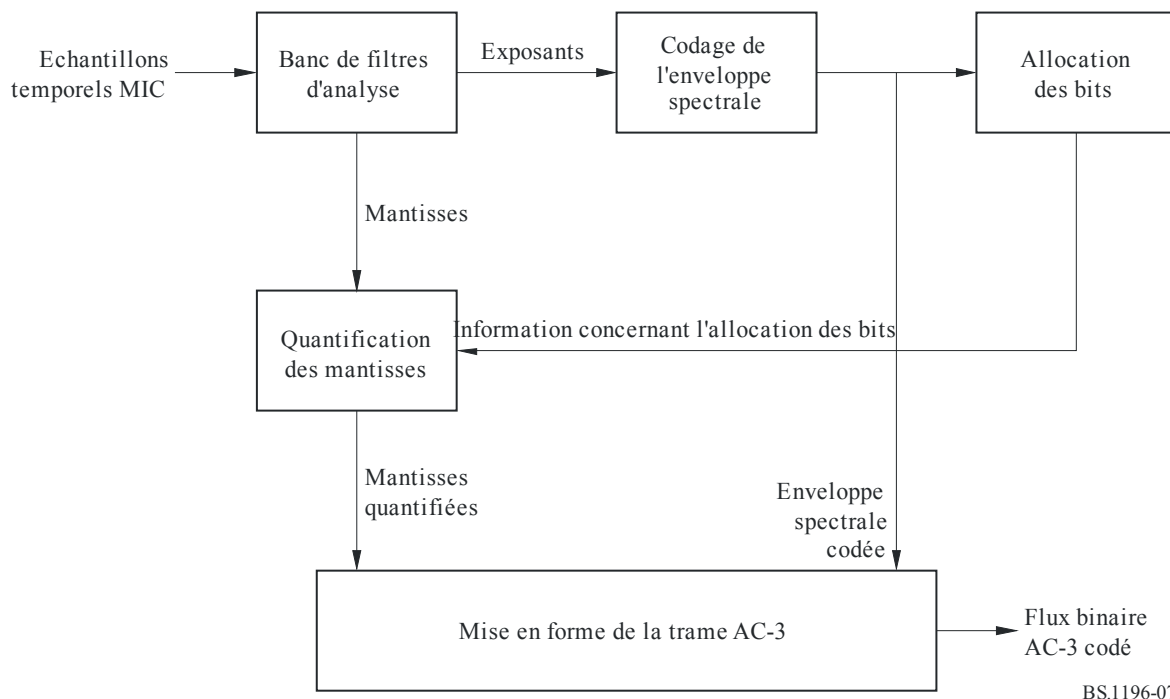
AC-3 et E-AC-3 audio

1 Codage

L'algorithme de compression numérique AC-3 peut coder des 1 à 5+1 canaux audio, à partir d'une représentation MIC, en un flux binaire série dont le débit est compris entre 32 et 640 kbit/s. L'algorithme AC-3 permet un gain de codage élevé (défini comme le rapport débit binaire en entrée sur débit binaire en sortie) par quantification approximative d'une représentation du signal audio dans le domaine fréquentiel. La Fig. 7 représente ce processus sous forme synoptique. La première étape du processus de codage consiste à transformer la représentation du signal audio, en partant d'une séquence d'échantillons temporels MIC pour obtenir une séquence de blocs de coefficients de

fréquence. Ceci est réalisé par le banc de filtres d'analyse. Des blocs de 512 échantillons temporels avec chevauchement sont multipliés par une fenêtre temporelle et subissent une transformation vers le domaine fréquentiel. En raison du recouvrement des blocs, chaque échantillon MIC entrant se trouve représenté dans deux blocs transformés consécutifs. Le domaine fréquentiel peut alors être réduit dans un rapport 2 de manière à ce que chaque bloc ne contienne que 256 coefficients fréquentiels. Chaque coefficient fréquentiel est représenté en notation exponentielle binaire sous la forme d'un exposant binaire et d'une mantisse. L'ensemble des exposants est codé sous la forme d'une représentation approximative du spectre du signal. Cette représentation, appelée enveloppe spectrale, est utilisée par le sous-programme de base d'allocation des bits, qui détermine le nombre de bits à utiliser pour coder chaque mantisse. L'enveloppe spectrale et les mantisses approximativement quantifiées de six blocs audio (1 536 échantillons audio) sont mis sous forme d'une trame AC-3. Le flux binaire AC-3 est une séquence de trames AC-3.

FIGURE 7
Codeur AC-3



En réalité, le fonctionnement du codeur AC-3 est plus complexe que ne l'indique la Fig. 7. En effet:

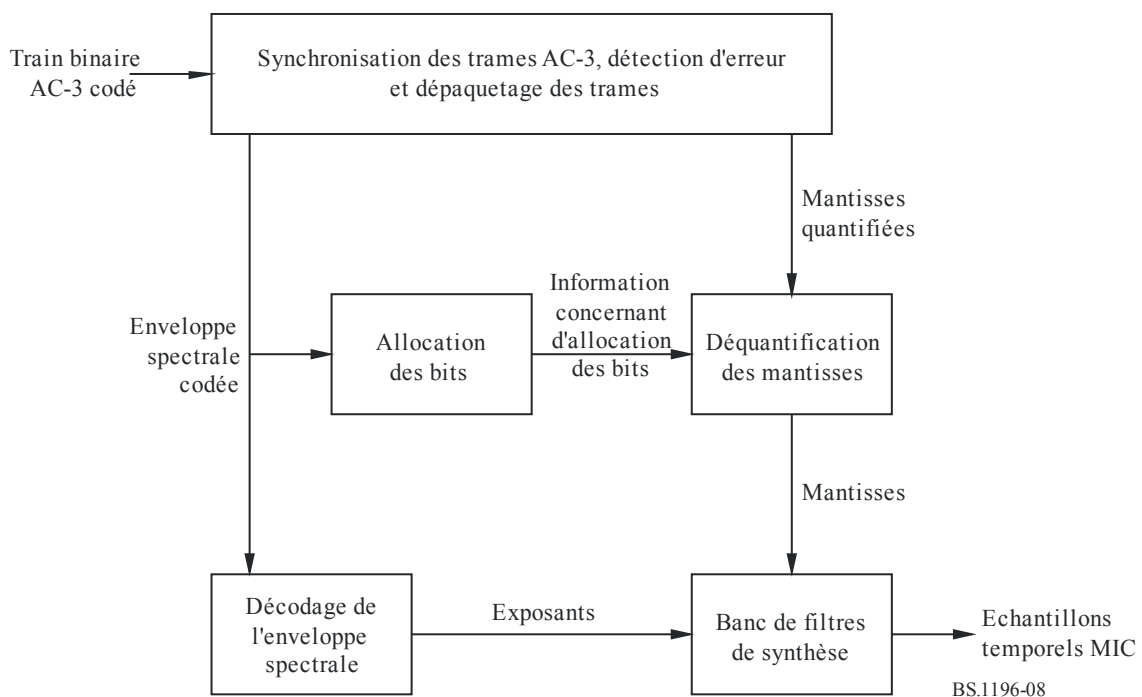
- un en-tête de trame est ajouté, contenant les informations nécessaires pour la synchronisation et le décodage du train binaire (débit binaire, fréquence d'échantillonnage, nombre de canaux codés, etc.);
- des codes de détection d'erreur sont insérés pour permettre au décodeur de vérifier qu'une trame reçue ne contient pas d'erreurs;
- la résolution spectrale du banc de filtres d'analyse peut être modifiée de façon dynamique pour assurer une meilleure adéquation des caractéristiques temps/fréquence de chaque bloc audio;
- l'enveloppe spectrale peut être codée avec une résolution temps/fréquence variable;
- une allocation des bits plus complexe peut être réalisée et les paramètres du sous-programme central d'allocation peuvent être modifiés à des fins d'optimisation;

- les canaux peuvent être couplés entre eux aux fréquences hautes en vue d'améliorer le gain de codage pour un fonctionnement à des débits binaires plus faibles;
- dans le mode bicanal, un processus de rematriçage peut être exécuté de manière sélective afin d'obtenir un gain de codage supplémentaire et d'améliorer les résultats dans le cas où le signal bicanal est décodé avec un décodeur ambiophonique matriciel.

2 Décodage

Le décodage est en substance le processus inverse du codage. Le décodeur présenté à la Fig. 8 doit se synchroniser sur le train binaire codé, détecter les erreurs et dépaqueter les différents types de données, notamment l'enveloppe spectrale codée et les mantisses quantifiées. Le sous-programme d'allocation des bits est exécuté et les résultats sont utilisés pour désassembler et déquantifier les mantisses et l'enveloppe spectrale est décodée de manière à produire les exposants. Les exposants et mantisses sont reconvertis dans le domaine temporel afin d'obtenir les échantillons temporels MIC décodés.

FIGURE 8
Décodeur AC-3



En réalité, le fonctionnement du décodeur AC-3 est plus complexe que ne l'indique la Fig. 8. En effet:

- en cas de détection d'erreur, le décodeur peut masquer l'erreur ou insérer un silence;
- les canaux dont les contenus à haute fréquence ont été couplés doivent être dégroupés;
- une opération de dématricage doit être exécutée (dans le mode bicanal) lorsque les canaux ont été rematricés;
- la résolution du banc de filtres de synthèse doit être modifiée de manière dynamique comme l'a été la résolution du banc de filtres d'analyse du codeur pendant le processus de codage.

3 E-AC-3

L'AC-3 amélioré (E-AC-3, *Enhanced AC-3*) ajoute plusieurs outils de codage et fonctions au codec AC-3 de base décrit ci-dessus. Les outils de codage supplémentaires améliorent l'efficacité du codage et permettent un fonctionnement à des débits binaires inférieurs, et les fonctions supplémentaires apportent une plus grande souplesse au niveau applicatif.

Outils de codage supplémentaires:

- transformée hybride adaptative – couche supplémentaire appliquée dans le banc de filtres d'analyse/de synthèse pour une résolution spectrale plus fine (1/6 de l'AC-3);
- traitement du pré-bruit transitoire – outil supplémentaire pour réduire le pré-bruit transitoire;
- extension spectrale – synthèse des composantes hautes fréquences par le décodeur reposant sur des informations auxiliaires créées par le codeur;
- couplage amélioré – gestion de la phase et de l'amplitude dans le couplage des canaux.

Fonctions supplémentaires:

- granularité du débit binaire plus fine;
- débit binaire maximum supérieur (3 Mbit/s);
- les sous-flux peuvent acheminer des canaux audio supplémentaires, par exemple des canaux 7.1, ou des pistes de commentaires.

Appendice 4

MPEG Surround

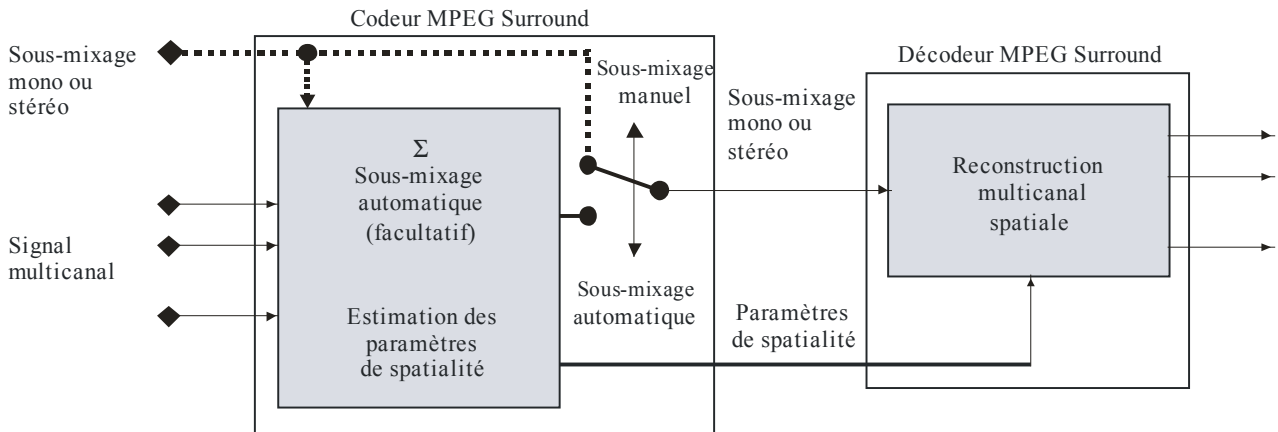
1 Introduction

La norme ISO/CEI 23003-1 ou technologie MPEG Surround est une méthode extrêmement efficace pour le codage du son multicanal, qui permet la transmission du son ambiophonique ou «*surround*» à des débits binaires couramment utilisés pour le codage du son mono ou stéréo. Cette technologie permet de représenter un signal audio multicanal à N canaux grâce à un sous-mixage (*downmix*) $M < N$ et des données de contrôle supplémentaires. Dans les modes de fonctionnement standard, un codeur MPEG Surround crée, à partir d'une entrée audio multicanal, un sous-mixage mono ou stéréo. Le sous-mixage obtenu est codé à l'aide d'un codec audio de base standard, par exemple l'un des systèmes de codage recommandés aux *recommande* 1 et 2. En plus du sous-mixage, le MPEG Surround produit une description paramétrique de l'image spatiale du signal audio multicanal. Cette description est ajoutée au codec audio de base sous la forme d'un flux de données auxiliaires, tout en assurant la rétrocompatibilité. Les décodeurs mono ou stéréo d'ancienne génération ne tiennent pas compte de ces données auxiliaires et lisent le sous-mixage mono ou stéréo. Les décodeurs compatibles MPEG Surround commencent par décoder le sous-mixage mono ou stéréo, puis utilisent les paramètres de description de l'image spatiale extraits du flux de données auxiliaires pour produire un signal audio multicanal de haute qualité.

La Fig. 9 illustre le principe du MPEG Surround.

FIGURE 9

Principe du MPEG Surround: le sous-mixage est codé à l'aide d'un codec audio de base



BS.1196-09

Grâce au MPEG Surround, les services existants peuvent être facilement mis à jour en vue d'offrir un son ambiophonique tout en restant rétrocompatibles. Alors que le décodeur stéréo d'un appareil d'ancienne génération ignore les données relatives au MPEG Surround et joue le signal stéréo sans dégradation de la qualité, un décodeur compatible MPEG Surround fournit un son multicanal de haute qualité.

2 Codage

Le codeur MPEG Surround a pour fonction de représenter un signal d'entrée multicanal sous forme de signal mono ou stéréo rétrocompatible, auquel sont associés des paramètres de spatialité permettant de reconstruire un signal de sortie multicanal qui ressemble aux signaux d'entrée d'origine du point de vue de la perception. Un sous-mixage extérieur (sous mixage dit «artistique») peut être utilisé à la place du sous-mixage produit automatiquement. Le sous-mixage doit préserver les caractéristiques de spatialité du signal sonore en entrée.

Le MPEG Surround repose sur la technologie stéréo paramétrique, qui, combinée au HE AAC, a donné la norme HE AAC v2. En combinant plusieurs modules de stéréo paramétrique et d'autres modules récemment développés, plusieurs structures prenant en charge des combinaisons différentes de nombre d'entrées et de canaux de sous-mixage ont été définies. Par exemple, pour un signal d'entrée multicanal 5.1, trois configurations sont possibles: une configuration pour les systèmes reposant sur le sous-mixage stéréo (configuration 525) et deux configurations différentes pour les systèmes reposant sur le sous-mixage mono (configurations 515₁ et 515₂ utilisant des cascades de modules différentes).

Le MPEG Surround intègre un certain nombre d'outils qui offrent des fonctions permettant une large application de la norme. L'une des fonctions essentielles est la capacité à faire varier la qualité de l'image spatiale de façon progressive, partant d'un surdébit spatial très faible jusqu'à une représentation du son transparente. Autre caractéristique importante, l'entrée du décodeur peut être rendue compatible avec les technologies ambiophoniques matricielles existantes.

Ces fonctions, ainsi que d'autres, sont assurées par les outils de codage suivants, qui occupent une place importante dans cette technologie:

- Codage résiduel: Outre les paramètres de spatialité, des signaux résiduels peuvent être acheminés à l'aide d'une technique de codage hybride. Ces signaux se substituent à une partie des signaux décorrélés (qui font partie des modules de stéréo paramétrique). Les signaux résiduels sont codés en passant du domaine QMF (filtres miroirs en quadrature) au

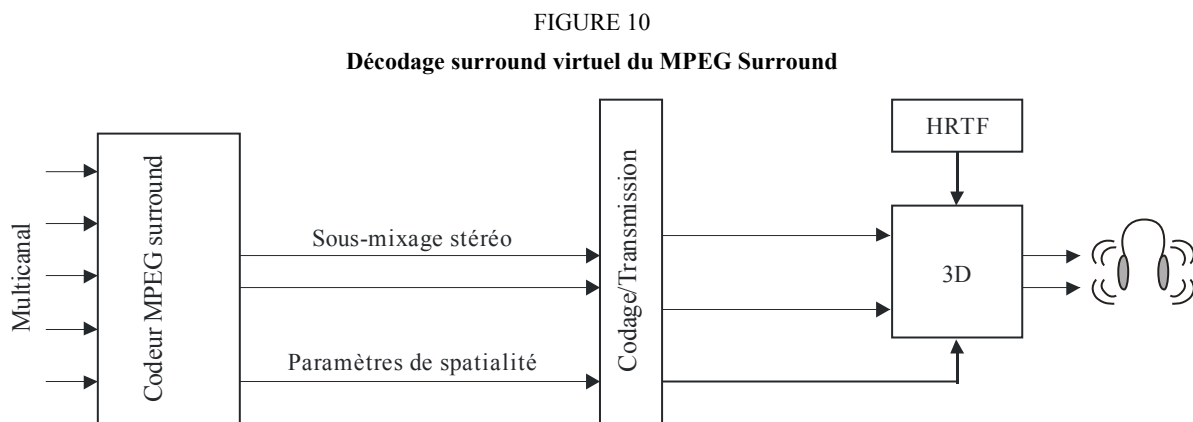
domaine TCDM (transformée en cosinus discrète modifiée) puis en codant les coefficients TCDM à l'aide de la norme AAC.

- Compatibilité matricielle: De manière facultative, le sous-mixage stéréo peut être prétraité pour le rendre compatible avec les technologies ambiophoniques matricielles d'ancienne génération en vue d'assurer une rétrocompatibilité avec les dispositifs qui ne peuvent décoder que le flux binaire stéréo, mais qui sont équipés d'un décodeur ambiophonique matriciel.
- Signaux de sous-mixage arbitraires: Le système MPEG Surround est capable de gérer, outre les sous-mixages générés par un codeur, les sous-mixages dits «artistiques» fournis au codeur en plus du signal multicanal d'origine.
- MPEG Surround sur MIC: En règle générale, les paramètres de spatialité du MPEG Surround sont acheminés dans la partie du système de compression audio sous-jacent réservée aux données auxiliaires. Dans le cas des applications pour lesquelles le sous-mixage est transmis au format MIC, le MPEG Surround prend également en charge une méthode permettant d'acheminer les paramètres de spatialité sur des canaux audio non compressés. La technologie sous-jacente est désignée sous le nom de «*buried data*» ou «données enfouies».

3 Décodage

Outre la restitution vers une sortie multicanal, le décodeur MPEG Surround prend en charge la restitution vers d'autres configurations de sortie:

- Surround virtuel: Le système MPEG Surround peut exploiter les paramètres de spatialité pour restituer le sous-mixage vers une sortie surround virtuelle stéréo connectée à un casque d'ancienne génération. La norme ne spécifie pas la fonction de transfert anatomique de l'individu (HRTF, *head-related transfer function*), mais seulement l'interface avec cette fonction, ce qui permet d'adapter la mise en œuvre au cas d'utilisation. Le traitement surround virtuel peut être appliqué à la fois dans le décodeur et dans le codeur, ce dernier offrant la possibilité d'un rendu ambiophonique virtuel sur le sous-mixage, sans qu'un décodeur MPEG Surround soit nécessaire. Le décodeur MPEG Surround peut toutefois annuler le traitement surround virtuel du sous-mixage et appliquer un autre traitement surround virtuel. Le principe de base est illustré à la Fig. 10.



- Mode matriciel amélioré: Dans le cas d'un signal stéréo d'ancienne génération, le MPEG Surround est capable d'estimer, à partir du sous-mixage, les informations auxiliaires de spatialité non contenues dans le signal et de créer ainsi une sortie multicanal offrant une qualité supérieure au système surround matriciel conventionnel.
- Elagage: Du fait de la structure sous-jacente, le décodeur MPEG Surround peut restituer sa sortie vers des configurations pour lesquelles les canaux sont en nombre inférieur à ceux présents dans l'entrée multicanal du codeur.

4 Profils et niveaux

Le décodeur MPEG Surround peut être mis en œuvre sous deux versions: l'une de qualité supérieure et l'autre de faible puissance. Bien que les signaux en sortie soient différents, les deux versions fonctionnent sur le même flux binaire.

Le profil de référence du MPEG Surround définit six niveaux hiérarchiques, qui diffèrent par le nombre de canaux en entrée et en sortie, par l'intervalle des fréquences d'échantillonnage et par la largeur de bande de décodage du signal résiduel. Pour garantir un décodage correct, le niveau du décodeur doit être supérieur ou égal au niveau du flux binaire. De plus, les décodeurs de niveau 1 et 2 sont capables de décoder tous les flux binaires de niveau 2 et 3, avec éventuellement une légère baisse de qualité due aux limites du décodeur. Par ailleurs, la qualité et le format de la sortie d'un décodeur MPEG Surround dépend de sa configuration. A noter que les aspects liés à la configuration du décodeur sont totalement indépendants des différents niveaux définis dans ce profil.

5 Interconnexion avec les codecs audio

Le MPEG Surround fonctionne comme une extension de pré- et post-traitement au-dessus des systèmes de codage audio d'ancienne génération. Il est donc capable de s'adapter à tout codeur audio de base, ou presque. Le MPEG Surround adopte une mise en trames très souple, qui s'adapte à une large gamme de codeurs. Il fournit en outre les moyens d'optimiser la connexion avec les codeurs utilisant déjà des outils paramétriques (par exemple, la reconstruction de bande spectrale).
